

Tavoiteoppimisen toteuttaminen tietokoneavusteisella arvioinnilla matematiikassa



HELSINGIN YLIOPISTO
MATEMATIIKAN JA TILASTOTIETEEN OSASTO

Timo Pelkola
Pro gradu -tutkielma

2019

Tiedekunta/Osasto — Fakultet/Sektion — Faculty		Laitos — Institution — Department	
Matemaattis-luonnontieteellinen		Matematiikan ja tilastotieteen osasto	
Tekijä — Författare — Author			
Timo Pelkola			
Työn nimi — Arbetets titel — Title			
Tavoiteoppimisen toteuttaminen tietokoneavusteisella arvioinnilla matematiikassa			
Oppiaine — Läroämne — Subject			
Matematiikka			
Työn laji — Arbetets art — Level		Aika — Datum — Month and year	
Pro gradu -tutkielma		Toukokuu 2019	
		Sivumäärä — Sidoantal — Number of pages	
		50 s.	
Tiivistelmä — Referat — Abstract			
<p>Tutkielmassa käsitellään tietokoneavusteisen arvioinnin hyödyntämistä tavoiteoppimisena (engl. mastery learning) tunnetun opetusmenetelmän yhteydessä.</p> <p>Tutkimuksen kohteena on Aalto-yliopiston Differentiaali- ja integraalilaskennan kursseilla toteutettu pilottihanke, jossa käytettiin STACK-arviointijärjestelmää tavoiteoppimismallin toteutukseen. Pilotoitu malli pohjautui Benjamin Bloomin Learning for Mastery -malliin, mutta Bloomin mallista poiketen tavoitetasoa mitattiin formatiivisten kokeiden lisäksi myös viikoittaisten tehtävien perusteella. Tavoiteoppimisen hengessä kurssiarviointiin kehitettiin myös uudenlainen tehtäväpisteiden laskentatapa, jossa pisteitä kertyi vasta tavoitetason ylittävistä tehtäväpisteistä. Kurssilla tavoiteoppimisen rooli oli erityisesti proseduraalisen osaamisen varmistaminen tietokoneavusteisesti, jonka vastapainoksi kursseille kehitettiin myös paperipohjaisia projektitehtäviä.</p> <p>Tutkimuskysymyksiä on kaksi: 1) soveltuuko STACK tavoiteoppimisen toteuttamiseen ja 2) pystyykö tavoitetason täyttymistä ennustamaan verkkotehtävien tuottaman datan perusteella?</p> <p>Ensimmäistä kysymystä tutkittiin toteutuksen onnistumisen, kuvailevien tilastojen ja kurssipalautteksen perusteella. STACK-järjestelmällä pystyi arvioimaan suurimman osan kurssin tavoitteista ja toteuttamaan opetusmenetelmän vaatimat formatiiviset kokeet. Kuvailevista tilastoista pystyi myös päättämään, että tavoitetasoon yltävien määrä kasvoi ensimmäisten yrityskertojen jälkeen (45% → 94%). Kurssipalautteksen perusteella formatiivisia kokeita myös pidettiin hyödyllisenä ja pisteytysmalli kannusti enemmän kuin lannisti tehtävien tekoon.</p> <p>Toista kysymystä varten sovellettiin ennustavaa mallintamista logistisella regressiolla. Näin saadun ennustemallin luokittelutarkkuus parani verrattuna siihen, että kaikkien oletetaan olevan ensiksi tavoitetason alapuolella (0.56 → 0.64). Tämä ei kuitenkaan riitä formatiivisten kokeiden korvaamiseen suoraan ennustemallilla. Ennustemallin tarkkuutta voidaan kuitenkin luultavasti parantaa.</p>			
Avainsanat — Nyckelord — Keywords			
Tavoiteoppiminen, tietokoneavusteinen arviointi, oppimisanalytiikka, STACK			
Säilytyspaikka — Förvaringsställe — Where deposited			
Kumpulan tiedekirjasto			
Muita tietoja — Övriga uppgifter — Additional information			

Sisältö

Johdanto	1
1 Käsitteet	4
1.1 Arviointi	4
1.1.1 Diagnostinen, formatiivinen ja summatiivinen	4
1.1.2 Arviointi matematiikassa	5
1.1.3 Tietokoneavusteinen arviointi	5
1.2 Tavoiteoppiminen	13
1.2.1 Teoreettinen tausta	13
1.2.2 Toteutus	16
1.2.3 Tavoiteoppimiseen liittyvä tutkimus	18
1.2.4 Kritiikki	20
1.2.5 Nykyinen käyttö	22
1.3 Oppimisanalytiikka ja opetuksen tiedonlouhinta	24
2 Tutkimus	26
2.1 Tausta	26
2.2 Tutkimusmentelmät	28
2.3 Tutkimuskysymykset	29
2.4 Tulokset	29
2.4.1 Kuvailevat tilastot	29
2.4.2 Kurssipalautekysely	31
2.4.3 Ennustemallit	33
2.5 Johtopäätökset	37
3 Pohdintaa	39
3.1 Bloomin tavoiteoppimisen kehittäminen	39
3.1.1 Tavoiteoppimismallien oletusten väljentäminen	39
3.2 Käytössä olevista arviointimenetelmistä	43

3.2.1	Harjoitustehtäväpisteet	43
3.2.2	Summatiivisesta arvioinnista formatiiviseen	44

Johdanto

Matematiikan tietokoneavusteinen arviointi on ollut aktiivisesti käytössä ja tutkimusaiheena 1980-luvulta lähtien, varhaisimpien kokeilujen ulottuessa 1950-luvulle. Edistysaskeleet web-teknologioissa, avoimen lähdekoodin ohjelmistoissa ja pääsyssä Internetiin ovat yleistäneet tietokoneavusteista arviointia merkittävästi. Tätä nykyä tietokoneavusteista arviointia hyödynnetään jossain määrin useimmissa suomalaisissa yliopistoissa, joissa monissa on käytössä tutkielmassa esiintyvä STACK-järjestelmä.

Automaattisen arvioinnin lisäksi nämä järjestelmät keräävät hyödynnettävissä olevaa dataa jokaisen opiskelijan oppimisen edistymisestä. Oppimisanalytiikka ja opetuksellinen tiedonlouhinta ovat aktiivisia ja suhteellisen uusia tutkimussuuntauksia, jotka yrittävät löytää tällaiselle tiedolle hyödyllistä käyttöä.

Tässä tutkielmassa pyritään havainnollistamaan miten tietokoneavusteista arviointia yhdessä sen tuottaman datan kanssa pystyisi hyödyntämään tavoiteoppimiseksi kutsutun opetusmenetelmän yhteydessä. Tavoiteoppimisen päätavoitteena on oppimiserojen kaventaminen. Tätä voi pitää perusteltuna ainakin teknisten alojen opintojen alkupään matematiikan kursseilla, joissa opittavat matematiikan taidot ovat usein esitietovaatimuksina myöhemmillä kursseilla, ja lopulta käytössä monissa tekniikan sovelluksissa.

Tutkielma koostuu keskeisten käsitteiden kirjallisuuskatsauksesta, tutkimuksen esittelystä ja pohdintaosiosta.

Tutkielmassa esitelty tutkimus (Pelkola, Rasila ja Christopher Sangwin 2018) suoritettiin osana Aalto Online Learning (A!OLE) -projektia kirjoittajan työskennellessä tutkimusapulaisena Aalto-yliopistolla. Tutkimuksessa suunniteltiin ja toteutettiin tietokoneavusteista tarkastamista hyödyntävä tavoiteoppimisen malli kahdelle pilottikurssille 2016-2017. Mallia arvioitiin järjestelmästä ja kurssipalautekyselyllä kerätyn datan pohjalta. Lisäksi tutkimuksessa kokeiltiin koneoppimisen menetelmiä tarkoituksena selvittää, voisiko ennustavalla mallintamisella parantaa pilotoitua mallia.

Luku 1

Käsitteet

1.1 Arviointi

Oppimisen arviointi voidaan määritellä prosessiksi, jonka tavoitteena on mitata oppimisen ja opetuksen laatua ja määrää käyttäen erilaisia tekniikoita, kuten tehtäviä, projekteja ja kokeita. (Page, Thomas ja Marshall 1979 lainattu teoksessa Miller, Imrie ja Cox 1998) Vaihtoehtoisesti arvioinnin voi määritellä todisteiden keräämiseksi joidenkin muutosten tapahtumisesta oppijoissa. (Bloom et al. 1971) Arvioinnilla viitataan myös arvioinnin lopputuotteeseen: "päätelmiin opiskelijan käyttäytymisestä sekä toisinaan saavutuksista ja affektiivisista tiloista". (E. L. Baker 2012)

Arviointia suoritetaan monista syistä, mm. oppimisen laadun osoittamiseksi opiskelijoille ja opettajille, laatutason ylläpitämiseksi ja opiskelijoiden motivoimiseksi. (Miller, Imrie ja Cox 1998) Jos arvioinnin tarkoituksena on sertifioida tai taata tason saavuttaminen, sen funktiona voidaan sanota olevan *ulosyöttö* (engl. feedout). Jos päämääränä on saattaa esiin oppimista hyödyntävää tietoa, arviointi toimii *palautteena* (engl. feedback). (Knight 2002) Tämä tutkielma keskittyy arvioinnin jälkimmäiseen funktioon.

1.1.1 Diagnostinen, formatiivinen ja summatiivinen

Arviointi voidaan jaotella *diagnostiseen*, *formatiiviseen* ja *summatiiviseen* sen käyttötavasta riippuen. (Bloom et al. 1971)

Opetusta edeltävää diagnostista arviointia käytetään oppijan sijoittamiseen sopivalle opetuksen tasolle. Diagnostinen arviointi voi selvittää esitietoja tai aikaisemman tason saavuttamista tai luokitella opiskelijoita kykyjen tai muiden piirteiden mukaan. Opetuksen aikana suoritettua diagnostista arviointia voidaan käyttää oppimisvaikeuksien taustalla olevien syiden kartoittamiseksi.

Formatiivinen arviointi suoritetaan opetuksen aikana ja se tarjoaa palautetta opiskelijoille ja opettajille. Formatiivista arviointia voidaan suorittaa hyvin epämuodollisesti esim. haastatteleamalla ja tarkkailemalla opiskelijoita (Fennell, Kobett ja Wray 2015) tai muodollisemmin oppimisosion päätteeksi formatiivisten kokeiden avulla.

Summatiivinen arviointi tapahtuu useimmiten kurssin tai oppimisosion lopuksi, ja sitä käytetään ensisijaisesti arvosanojen määrittelymiseen. Summatiivinen arviointi voi olla myös *jatkuvaa*, jos oppimisen edistymistä mitataan ja raportoidaan pitkin kurssia.

Näiden luokkien välillä on myös jonkin verran päällekkäisyyksiä. Esimerkiksi Taras (2005) esittää, että vaikka arviointi voi olla ainoastaan summatiivisesta, se ei voi olla ainoastaan formatiivista ilman sitä edeltänyttä summatiivista arvostelua. Jatkovaa summatiivista arviointia voidaan taas käyttää myös formatiivisen arvioinnin tavoin.

1.1.2 Arviointi matematiikassa

Matematiikassa arvioidaan yleensä tarkastelemalla opiskelijan vastausta annettuun tehtävään, josta tehdään päätelmiä opiskelijan matemaattisesta osaamisesta.

Ylioppilastutkintolautakunnan (2016) mukaan hyvä vastaus näyttää miten opiskelija on päätenyt lopputulokseen ja sisältää tarvittavat laskut ja muuten riittävät perustelut. Vastausta arvioidaan kolmessa osassa: alku, välivaiheet ja lopullinen vastaus. Pienistä laskuvirheistä sakotetaan vähemmän kuin tehtävänantoa oleellisesti muuttavista virheistä.

On huomattava, että arvioinnin tarkoitus on mitata oppimisen laatua ja määrää eikä niinkään vastauksen, jonka oletetaan olevan oppimisen indikaattori. Smith et al. (1996) esittelee muokatun version Bloomin taksonomiasta matemaattisen tehtävien luokitteliseksi sen mukaan, mitä tehtävien on tarkoitus mitata. (Ks. Taulukko 1.1).

1.1.3 Tietokoneavusteinen arviointi

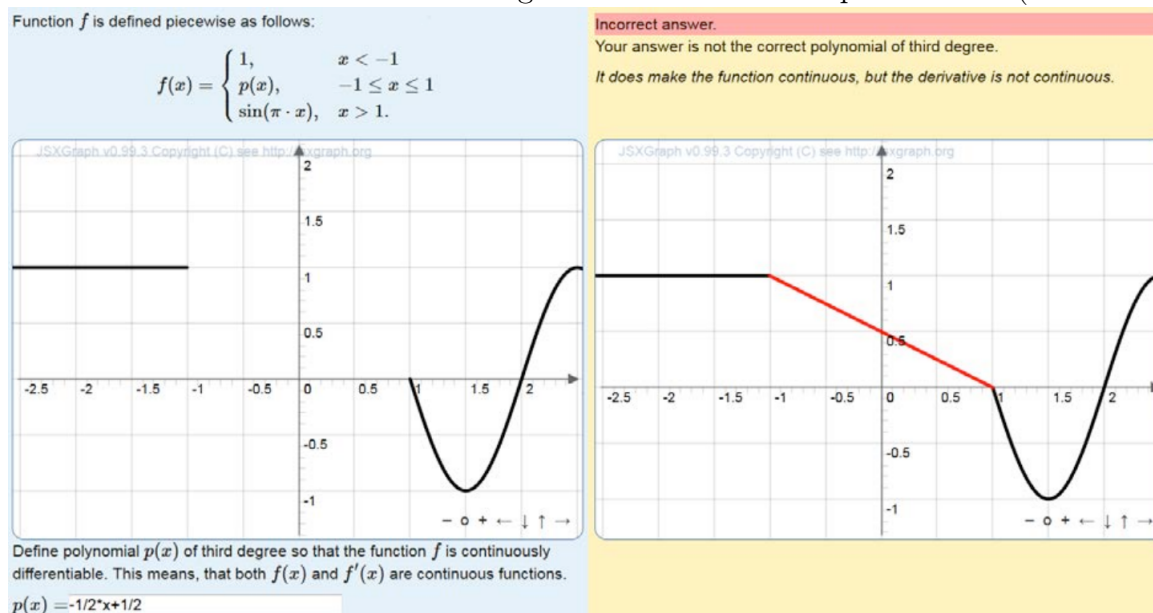
Tietokoneavusteinen arviointi voidaan määritellä tietokoneiden käyttönä tehtävien tai kokeiden jakamiseen, pisteytykseen ja analysointiin. Tietokoneavusteista arviointia käytetään monista syistä, esim. opiskelijoiden motivoimiseksi, pisteytysurakan keventämiseksi tai palautteen lisäämiseksi. (Bull ja McKenna 2003)

Arviointijärjestelmän antaman palautteen laatu voi vaihdella niukasta (oikein/väärin) kohtuulliseen (mallivastaus) ja rikkaaseen (vihjeet ja kommentit). (Trenholm, Alcock ja Robinson 2015) Rikas palaute voi liittyä suoraan opiskelijan tekemään virheeseen. Arviointijärjestelmä voi esim. kertoa, jos opiskelija on derivoinnin sijaan vahingossa integroinut, tai jos vastaus on osittain oikein (kuva 1.1). Jokaisen mahdollisen virheellisen vastauksen luokittelu on kuitenkin epätriviaalia, varsinkin automaattisesti (asiaa on käsitelty lisää esim. tutkielmassa Tiitu (2016)).

Taulukko 1.1: MATH (Mathematical Assessment Task Hierarchy) -taksonomia tiivistettynä (Smith et al. 1996). Esimerkit artikkelista Smith et al. (1996), paitsi "sovellukset uusissa tilanteissa" (Jungic, Menz ja Pyke 2011) ja "proseduurien rutiinikäyttö" (kirjoittaja).

Ryhmä A	Kyky:	Esimerkki
<i>Faktuaalinen tieto</i>	palauttaa mieleen tietoa annetussa muodossaan	<i>Mitä tarkoitetaan käsitteellä lineaarinen differentiaaliyhtälö?</i>
<i>Ymmärtäminen</i>	päätellä milloin <i>yksinkertaisen</i> määritelmän ehdot täyttyvät, ymmärtää symbolien merkitykset kaavassa, tunnistaa esimerkkejä	<i>Osoita, että $x^3 - 3xy^2$ on harmoninen funktio.</i>
<i>Proseduurien rutiinikäyttö</i>	käyttää harjoiteltuja proseduureja oikein	<i>Ratkaise $x: x^2 - 5x + 6 = 0$</i>
Ryhmä B		
<i>Tiedon muokkaaminen</i>	muokata tietoa yhdestä muodosta toiseen, päätellä milloin <i>konseptuaalisen</i> määritelmän ehdot täyttyvät, tunnistaa kaavan tai menetelmän soveltuvuus, tiivistää ja selittää	<i>Osoita, että yhtälöllä $x^3 + cx + d = 0$ on vain yksi reaalinen juuri, jos $c \geq 0$.</i>
<i>Soveltaminen uusissa tilanteissa</i>	valita ja soveltaa sopiva menetelmä tai tieto uusissa tilanteissa, kuten todellisen ilmiön mallintamisessa tai ennalta tuntemattoman lauseen todistuksessa, joka ylittää proseduurien rutiinikäytön	<i>Kaksi ravihevosta aloittaa juoksun samanaikaisesti ja saapuu maaliin samalla hetkellä. Osoita, että jossakin vaiheessa ravia hevoset juoksee samaa vauhtia.</i>
Ryhmä C		
<i>Perusteleminen ja tulkitseminen</i>	perustella ja/tai tulkita lopputulos, esim. todistamalla lause tai tunnistamalla mallin rajoitukset	<i>Eräs tapa mallintaa väestönkasvua on olettaa, että väestönkasvun nopeus on verrannollinen väestön kokoon, ts. että $P' \propto P$. Milloin tällainen malli sopii väestönkasvun mallintamiseen? Mitä oletuksia mallissa tehdään ja mitkä on sen rajoitukset?</i>
<i>Seurausten, otaksu- mien ja vertailujen tekeminen</i>	päätellä annetun tuloksen seuraukset, tehdä ja todistaa otaksunia induktiivisten tai heurististen perusteluiden pohjalta, vertailla algoritmeja, konstruoida esimerkki	<i>Todista oikeaksi tai vääräksi seuraava väittämä: lauseke $n^2 + n + 41$ esittää luonnollista lukua kaikilla luonnollisilla luvuilla n.</i>
<i>Arviointi</i>	tehdä arviointeja, valita oleellisuuden perusteella, olla luova	<i>Kirjoita lyhyesti Leibnizin ja Newtonin derivaattanotaatioiden suhteellisista eduista.</i>

Kuva 1.1: Esimerkki STACK-tehtävän generoimasta rikkaasta palautteesta (Rasila 2016)



Monivalintatehtävät

Monivalintatehtävät eri muodot mukaan lukien (esim. monivaste, tosi/epätosi) on yleinen tehtävätyyppi, jonka tarkastaminen on helposti automatisoitavissa. Oikean ratkaisun valitsemisen ja luomisen välinen ero tekee kuitenkin hyödyllisten matematiikan monivalintatehtävien tekemisestä hankalaa (Chris Sangwin 2013), kuten seuraavat esimerkit osoittavat:

1. Ratkaise x

$$x^2 - 5x + 6 = 0$$

- a) $x = 2$ tai $x = 3$ b) $x = 1$ tai $x = -1$ c) $x = 1$ tai $x = 3$

2. Määritä

$$\int 3x^2 \, dx$$

- a) $6x + C$ b) $6x^3 + C$ c) $x^3 + C$

Vaikka melko varmasti voi sanoa, että väärän vastausvaihtoehdon valinnut opiskelija ei osaa ratkaista toiseen asteen yhtälöä tai integroida polynomilauseketta, näitä ei tarvitse osata myöskään oikean vastausvaihtoehdon valitsemiseen. Oikean vastauksen voi tarkistaa sijoittamalla tehtävässä 1 ja derivoimalla tehtävässä 2. Näin ollen tehtävät eivät ole luotettavia mittareita sille mitä ne vaikuttavat mittaavan.

Oikein arvaamisen mahdollisuus täytyy ottaa myös huomioon monivalintakokeissa. Koetehtävien tai vastausvaihtoehtojen määrän kasvattaminen pienentää todennäköisyyttä saada korkeita pistemääriä satunnaisesti arvaamalla. Sopivien *harhauttajien* eli väärin vaihtoehtojen keksiminen on kuitenkin yleensä vaikeaa ja tehtävien määrää rajoittaa myös käytettävissä oleva koeaika ja muut asiat. Zhao (2006) esittää käytännölliseksi vastausvaihtoehtojen lukumääräksi neljää ja asteikkoja, joissa arvaamisen oletettu vaikutus on otettu huomioon.

Arvaamisen rajoittamiseksi ja kompensoimiseksi on olemassa myös erilaisia pistesakkomalleja ja tapoja ottaa vastausvarmuus huomioon. Näissäkin ongelmana voi olla, että jotkut opiskelijat ovat enemmän taipuvaisia ottamaan riskejä kuin toiset, mikä saattaa vääristää koetuloksia. (Sherriffs ja Boomer 1954)

Toinen huoli on, että monivalintatehtävät saattavat johtaa harhauttajien muistamiseen oikean vastauksen sijaan. (Chris Sangwin 2013; Roediger III ja Marsh 2005) Monivalintakokeet saattavat myös pahimmillaan kannustaa opiskelijoita pintasuuntautuneeseen oppimistapaan. (Scouller 1998)

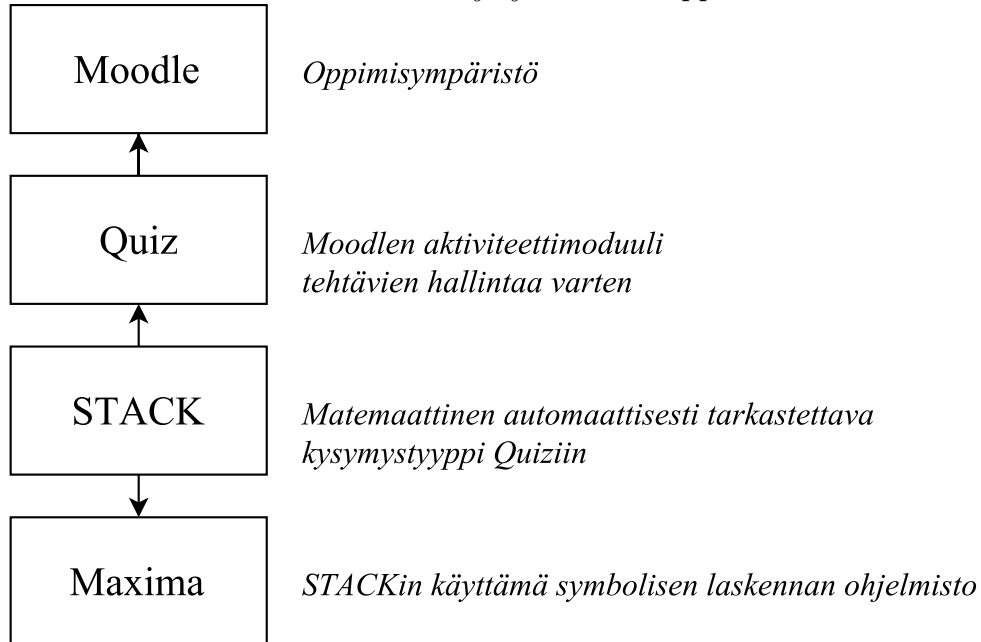
STACK

Osalta edellä mainituista monivalintatehtäviin liittyvistä ongelmista voidaan välttyä varustamalla arviointijärjestelmä symbolisella laskimella. Symbolinen laskin mahdollistaa matemaattisen lausekkeen syöttämisen vastauksena, jonka jälkeen vastauksen matemaattisia ominaisuuksia voidaan tarkastella oikeellisuuden arvioimiseksi.

STACK on Chris Sangwinin (2013) johtama avoimen lähdekoodin projekti, jonka taustalla on Maxima-nimistä symbolisen laskennan ohjelmisto. STACK on suunniteltu:

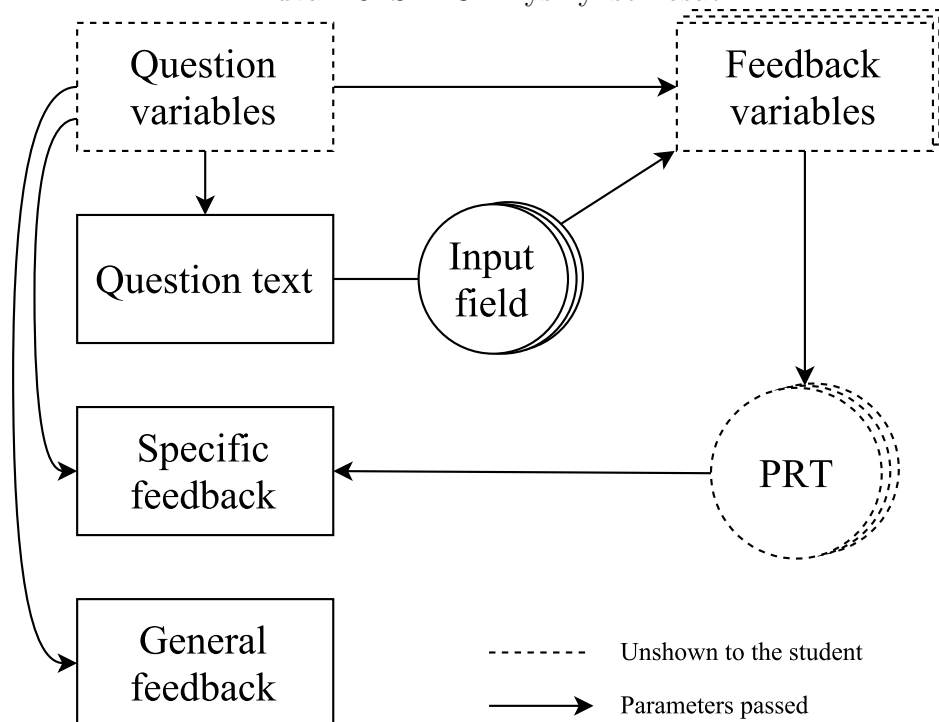
- generoimaan satunnaistettuja versioita tehtävistä strukturoidulla matemaattisella tavalla
- hyväksymään matemaattista sisältöä sisältäviä opiskelijoiden vastauksia monivalintatehtävien sijaan
- tunnistamaan opiskelijoiden vastausten matemaattiset ominaisuudet
- toteuttamaan formatiivista, summatiivista ja opetuksen arviointia
- tallentamaan dataa opiskelijan vastausyrityksistä tehtävän tai opiskelijan mukaan opettajan analyysia varten

Kuva 1.2: STACK-järjestelmän riippuvuudet



STACK-kysymys koostuu *tehtävämuuttujista*, *tehtävätekstistä*, opiskelijan syötteen mahdollistavista *interaktioelementeistä* ja yhdestä tai useammasta *vastauspuuksi* kutsutusta tarkastusalgoritmista, joita kutakin varten voi määritellä *palautemuuttujia*. Vastauspuut generoivat *erityisen palautteen*, mutta tehtävään voi liittyä myös *yleinen palaute* (esim. mallivastaus).

Kuva 1.3: STACK-kysmyksen osat



Tehtävämuuttujat määritellään Maxima-koodilohkossa ja niiden arvoihin pääsee käsiiksi muista tehtävän osista. Tehtävämuuttujia käytetään parametrien satunnaistamiseen ja muihin tehtävän kannalta hyödyllisiin laskuihin. Esim. yksi tapa konstruoida polynomi, jolla on satunnaiset juuret $a, b \in \{-3, -2, \dots, 2, 3\}$ olisi avata lausekkeen $(x - a)(x - b)$ sulkeet kirjoittamalla:

```

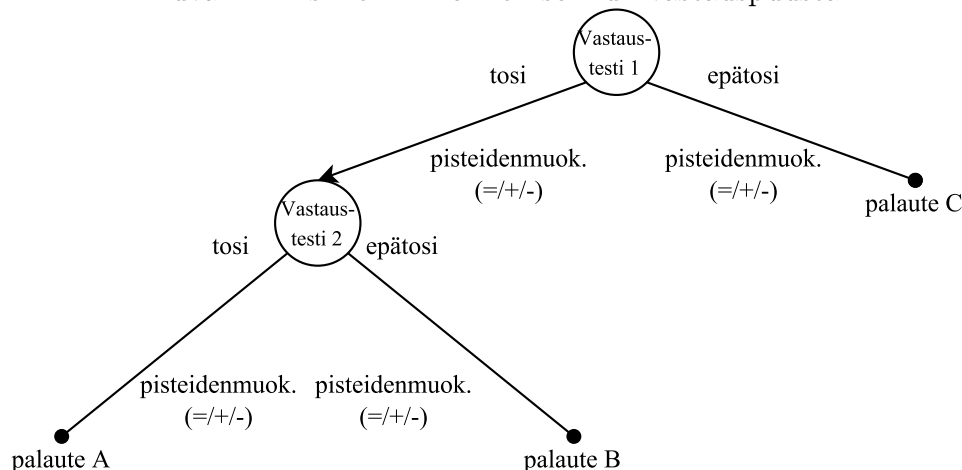
a : rand ([ -3 , -2 , -1 , 0 , 1 , 2 , 3 ])
b : rand ([ -3 , -2 , -1 , 0 , 1 , 2 , 3 ])
poly : expand (( x-a )*( x-b ))
    
```

Tehtäväteksti sisältää varsinaisen tehtävänannon \LaTeX ja HTML-merkkintäkielillä muotoiltuna sekä kaikki opiskelijoiden käyttämät interaktioelementit, kuten tekstikentät ja vetolaatikot. Maximaa voidaan käyttää myös tehtävätekstin seassa tehtävämuuttujien arvojen hakemiseen, laskutoimitusten tekemiseen tai ladontaan.

Opiskelijoiden vastaukset arvioidaan kulkemalla vastauspuu läpi. Jokaisen puun solmun kohdalla opiskelijan vastauksen sisältämä lauseke saa totuusarvon tosi tai epätosi. Tuloksesta riippuen algoritmi voi antaa tai vähentää pisteitä, generoida palautetta tai siirtyä seuraavaan solmuun. Lausekkeet kirjoitetaan valitsemalla sopiva *vastaustesti*, joka vertaa opiskelijan vastausta opettajan vastaukseen esim. ekvivalenssin, tarkkuuden

tai muodon kannalta. Interaktioelementit ja vastauspuut ovat erillisiä, mikä mahdollistaa monia interaktioelementtejä sisältävän tehtävän tarkastamisen joustavasti.

Kuva 1.4: Esimerkki kolmen solmun vastauspuusta



STACK on ollut käytössä, kehitettävänä ja tutkittavana (ks. Matta 2017) Aalto-yliopistossa¹ vuodesta 2006. Aallon koordinoima lähinnä STACK-kysymyksistä koostuva materiaalipankki Abacus (2016) perustettiin vuonna 2015 yhteistyössä kuuden muun suomalaisen yliopiston kanssa. Myöhemmin mukaan on tullut muita yhteistyökumppaneita useista eri maista.

Tietokone- ja ihmisarvioijan vertailua

Matemaattinen notaatio on perinteisesti kirjoitettu käsin tai ladottu luettavuuden ehdoilla. Tämä on johtanut notaatioon, joka sisältää paljon symboleita, kirjoitetaan kaksiulotteisesti ja on riittävän formaalia ihmisen luettavaksi.

Tällaista notaatiota on yleisesti vaikea tuottaa ja käsitellä tietokoneilla. Merkintäkielekset, kaavaeditorit ja käsialatunnistus saattavat auttaa toisintamaan esitystavan, mutta ei notaation matemaattista merkitystä. Usein tietokone tarvitsee vähemmän tulkinnanvaraisen syntaksin (esim. Maximasta löytyvä) pystyäkseen tulkitsemaan matemaattisen syötteen täsmällisesti.

Uuden syntaksin opetteleminen vaatii jonkin verran vaivannäköä, ja puuttuvan kerto-merkin tai sulkeiden tyyppiset syntaksivirheet ovat tavallisia. Jotkin matemaattiset konstruktiot, kuten todistukset, ovat erityisen vaikeita ilmaista täsmällisellä formaalilla syntaksilla.

¹Teknillinen korkeakoulu ennen vuotta 2010

Matemaattisen notaation syöttämisen ja tulkitsemisen hankaluus on johtanut taipumukseen tarkastaa tietokoneavusteisesti vain osia opiskelijan ratkaisusta, esim. lopullinen vastaus. Tämä kuitenkin rajoittaa myös annettavan palautteen syvällisyyttä, koska lopullisesta vastauksesta on vaikeata päätellä minkä virheen opiskelija on mahdollisesti tehnyt. Välivaiheiden arviointi on mahdollista rajoitetusti valmiiksi annetuilla vastauskentillä, mutta tämä voi myös muokata arvioitavaa tehtävää merkittäväällä tavalla esim. helpottaen tai poissulkien vaihtoehtoisia ratkaisumenetelmiä.

Kuva 1.5: Välivaiheellinen tehtävä

Integrate

$$\int_0^1 x^2 \sqrt{x^3 + 7} \, dx$$

using the substitution $u = x^3 + 7$.

$$du = \boxed{} \, dx$$

The new lower limit:

The new upper limit:

$$\int_0^1 x^2 \sqrt{x^3 + 7} \, dx = \boxed{}$$

Tietyntyypisten lopputulokseltaan avointen ongelmatehtävien (engl. open-ended problems), kuten "anna esimerkki polynomifunktioista, joiden derivaatat kohdassa $x = 0$ saa arvon 4", tarkastaminen tietokoneavusteisesti on mahdollista. Tällaisen tehtävän jokainen oikea vastaus on ennakoitavissa, hyvin määritelty ja tarkastettavaa tyyppiä. Kuten Yeo (2007) toteaa, ongelmatehtävän avoimuuden voi nähdä jatkumona. Mitä avoimempi tehtävä on, sen monimutkaisempaa sen arviointi yleisesti ottaen on. Jos opiskelija pystyy antamaan vain lopullisen vastauksen, tämä tekee ratkaisumenetelmiltään avointen ongelmatehtävien (engl. open-method problem) arvioimisesta mahdotonta.

Avointen ongelmatehtävien vastausten arviointiin liittyy myös joitain näkökulmia, joita suljetuissa tehtävissä ei tule esiin. (Nohda 2000) mainitsee näistä:

- *sujuvuuden* eli montako ratkaisua opiskelija keksii,
- *joustavuuden* eli montako erilaista matemaattista ideaa opiskelija löytää
- *omaperäisyyden* ja
- *tyylikkyyden* eli kuinka yksinkertainen ja selkeä opiskelijan ratkaisun idea on?

Tällaisten näkökulmien sisällyttäminen arviointialgoritmiin paitsi tekisi algoritmista monimutkaisen toteuttaa vaatisi myös, että kriteerit näiden täyttymiselle tulisi olla ennalta määriteltynä.

1.2 Tavoiteoppiminen

Tavoiteoppiminen (engl. mastery learning) on opetusfilosofia ja opetusmenetelmien joukko, joita tunnetuin Benjamin Bloom (1968) kehitti eteenpäin. Tavoiteoppimisen filosofia lähtee ajatuksesta, että sopivissa oppimisolosuhteissa suurin osa opiskelijoista saavuttaisi korkean oppimisen tason, kun tavallisesti sen koulussa saavuttaa vain pieni osa. Tavoiteoppimisen menetelmät korostavat formatiivista arviointia ja kriteereihin perustuvan tavoitetason saavuttamista ennen seuraavaan oppimisosioon siirtymistä.

Taitavan opettajan antamaa yksityisopetusta voisi pitää ideaalina oppimisolosuhteena. Kun opiskelija tekee yksityisopetuksessa jonkin virheen, hän saa välitöntä palautetta ja apua asian ymmärtämiseksi. Eräs Bloomin havainnoista oli, että keskiverto-opiskelija saa yksityisopetuksessa kahden keskihajonnan verran parempia koetuloksia perinteiseen luokahuoneopetukseen verrattuna (ts. suoriutuu paremmin kuin 98% opiskelijoista luokahuoneopetuksessa). (Bloom 1984) Bloom nimesi vastaaviin tuloksiin ilman yksityisopetusta pääsemisen *kahden sigman ongelmaksi*.

1.2.1 Teoreettinen tausta

Carrollin kouluoppimisen malli

Bloomin tavoiteoppimiseen vaikutti merkittävästi John Carrollin (1989) kehittämä kouluoppimisen malli, jossa koulumenestystä selittää viisi muuttujaa.

Kolmea muuttujista voi kuvata ajan avulla:

- *Kyvykkyyden* määrittelee aika, jonka opiskelija **tarvitsee** oppiakseen annetun asian hyväksyttävien kriteerien mukaisella tasolla ja opetusolosuhteiden ja motivaation ollessa optimaalisia. Korkeampaa kyvykkyyttä kuvaa lyhyempi aika.

Bloom ajatteli tämän muotoilun sisältävän perusteellisia seurauksia opetukselle, sillä siinä oletetaan implisiittisesti, että kunhan aikaa annetaan riittävästi, kaikki opiskelijat voivat päästä tavoitetasolle. Bloomin mukaan koulumenestyksen ennustaminen kyvykkyyttä mittaavilla testeillä oli johtanut käsitykseen, jossa vain parhaimmat opiskelijat voivat saavuttaa korkean oppimisen tason.

Bloomin käsityksen mukaan kyvykkyys ennusti oppimistahtia eikä mahdollista oppimisen tasoa, ainakin kyvykkyydeltään keskimmäselle 90 prosentille. Bloom kuitenkin tunnusti, että jakauman ääripäissä on luultavasti erityislahjakkaita ja toisaalta oppimisvaikeuksisia opiskelijoita, jotka voivat yltää muita korkeammalle tai vain matalammalle tasolle. (Bloom 1968)

- *Oppimismahdollisuudet* määritellään aikana, joka **annetaan** oppimiselle.
- *Sinnikkyys* määritellään aikana, jonka opiskelija on **halukas käyttämään** annetun tehtävän oppimiseen.

Lisäksi *oppimisen laadun ja kyvyn ymmärtää opetusta* oletetaan olevan yhteyksissä aikaan, joka tarvitaan oppimiseen.

Näin oppimisen tason (L) voi ajatella kasvavana funktiona, jonka muuttujana on oppimiseen käytetyn ja oppimiseen vaaditun ajan välinen suhde:

$$L = f\left(\frac{\text{käytetty aika}}{\text{tarvittu aika}}\right),$$

jossa käytetty aika on pienin tarvitusta ajasta, oppimismahdollisuuksista ja sinnikkyydestä.

Tärkeä seuraus tälle muotoilulle on, että oppimista voidaan parantaa kasvattamalla oppimiseen käytettyä aikaa tai vähentämällä oppimiseen tarvittavaa aikaa. Mallin mukaan kaikki saavuttavat tavoitetason riittävän ajan kuluessa.

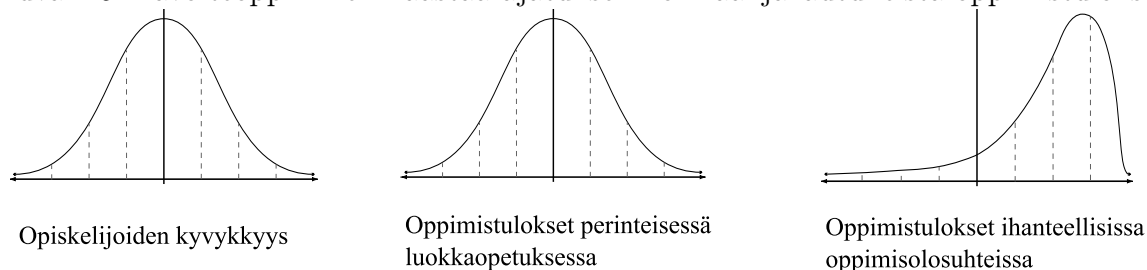
Bloomin tulkitsi, että jos kaikki opiskelijat ovat normaalijakautuneita kyvykkyyden suhteen ja saavat ajallisesti ja laadullisesti täsmälleen samaa opetusta, saavutukset ovat silloin myös normaalijakautuneita. Jos taas opetusta ja oppimiseen annettua aikaa mukautetaan jokaiselle oppijalle sopivaksi, suurin osa saavuttaa tavoitetason. (Bloom 1968)

Bloomin kouluoppimisen malli

Bloom (1976) kehitti myöhemmin omaa teoriaansa tavoiteoppimisen tutkimuksen pohjalta, joka osin muistuttaa Carrollin teoriaa, mutta ei keskity samalla tavalla ajankäyttöön.

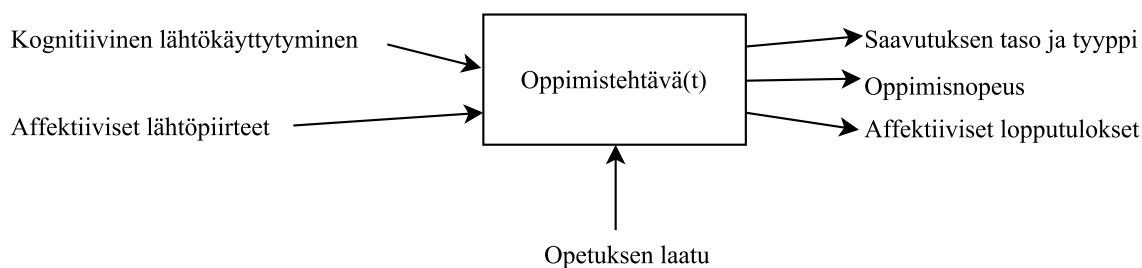
Bloomin teoriassa tarkisteltava yksikkö on yksittäinen *oppimistehtävä* (engl. learning task) ja oppijan historialle annetaan paljon painoarvoa. Kouluoppimisen vaihtelua selitetään kolmella toisistaan riippuvilla muuttujilla:

Kuva 1.6: Tavoiteoppiminen haastaa ajatuksen normaalijakautuneista oppimistuloksista



Kuva 1.7: Tärkeimmät muuttujat Bloomin kouluoppimisen teoriassa

OPISKELIJAN PIIRTEET	OPETUS	OPPIMISEN LOPPUTULOKSET
----------------------	--------	-------------------------



- *Kognitiivinen lähtökäyttäytyminen* (engl. cognitive entry behaviours) eli oppimistehtävän vaatima edeltävä oppiminen
- *Affektiiviset lähtöpiirteet* (engl. affective entry characteristics) eli opiskelijan motivaatio oppia oppimistehtävä
- *Opetuksen laatu* (engl. quality of instruction) eli kuinka sopivia annetut informaatio, harjoitus ja vahvistaminen ovat oppijan tarpeisiin nähden

Oppimisen lopputulos voidaan taas jakaa edelleen kolmeen osaan:

- *Saavutuksen taso ja tyyppi*
- *Oppimismnopeus*
- *Affektiiviset lopputulokset*

Bloom kollegoineen hylkäsi ajatuksen "hyvistä ja huonoista oppijoista", mutta lopulta myös ajatuksen "nopeista ja hitaista oppijoista", joka seurasi Carrollin mallista. Heidän

tekemänsä tutkimus saivat heidät vakuuttuneiksi, että "suurin osa opiskelijoista rupeaa muistuttamaan toisiaan hyvin paljon oppimiskykyjen, oppimismenopeuden ja motivaation oppia uutta kannalta, kun heille annetaan suotuisat oppimisolosuhteet".

1.2.2 Toteutus

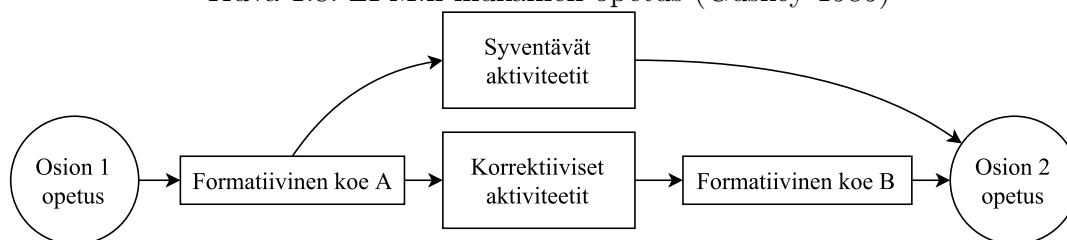
Tavoiteoppimista tutkinut Thomas Guskey (1986) pitää *palaute* ja *korrektiiveja* ja *yhtenevyyttä opetuksen komponenttien välillä* tärkeimpinä tavoiteoppimistoteutuksen elementteinä. Tavoiteoppimisessa opiskelijoille tulisi antaa säännöllisesti tietoa oppimisen edistymisestä eli palautetta. Palautteen tulisi olla sekä diagnostista että ohjaavaa ja kertoa opiskelijoille mitä on tärkeää oppia, mitä he ovat jo oppineet ja mitä heidän täytyisi opetella vielä lisää.

Palautteen mukana pitäisi aina antaa korrektiiveja eli täsmällistä ohjausta, jonka avulla oppija voi lähteä korjaamaan oppimistaan. Korrektiivit ihanteellisesti eroavat alkuperäisestä opetuksesta, jos aikaisempi lähestymistapa on jo osoittautunut epäonnistuneeksi.

Opetuksen komponentit sisältävät *oppimistavoitteet*, *opetuksen*, *palautteen* ja *korrektiivit* sekä *osaavat oppijat*. Yhtenevyys näiden välillä tarkoittaa, että opetus, palaute ja korrektiivit ja oppimisen arviointi vastaavat sitä, mitä oppijoiden oletetaan oppivan.

Ryhmämuotoinen Learning for Mastery

Kuva 1.8: LFM:n mukainen opetus (Guskey 1985)



Vaikka Bloom kirjoittikin, että "tavoiteoppimiselle on olemassa monia vaihtoehtoisia strategioita", useimmiten tavoiteoppimisella viitataan ryhmämuotoiseen *Learning for Mastery* -malliin (LFM). Tässä yhteydessä ryhmämuotoisuudella tarkoitetaan, että opetus suunnataan koko ryhmälle ja on opettajajohtoisesti ajoitettua. Tämä on perinteisesti tehnyt LFM:n sovittamisesta kouluympäristöön kohtalaisen helppoa.

LFM:ssä kurssin sisältö jaetaan ensiksi *oppimisosioihin* (engl. learning unit), jotka vastaavat laajuudeltaan esim. oppikirjan lukua tai 1-10 tunnin kestävästä opiskelusta. *Oppimistavoitteet* määritellään oppimisosioittain ja esitetään selkeästi opiskelijoille.

Ryhmämuotoisen oppimisosion opetuksen jälkeen opiskelijoille annetaan formatiiviset kokeet, joilla selvitetään onko opiskelija saavuttanut tavoitetason. Tavoitetaso määritellään yleensä osuutena koepisteistä ja se liikkuu tyypillisesti 80-95% välillä.

Tavoitetason saavuttamattomille opiskelijoille annetaan korrektiiveja, jotka sisältävät yleensä vaihtoehtoisia oppimateriaaleja ja oppimismenetelmiä. Korrektiivien tulisi kohdistua juuri formatiivisissa kokeessa ilmi tulleisiin heikkouksiin.

Taulukko 1.2: Guskeyn (1985) esittämiä korrektiiveja

Korrektiivi	Opettajan kanssa	Toverin kanssa	Itsekseen
Uudelleenopetus	✓		
Kurssin oppikirja	✓	✓	✓
Vaihtoehtoiset oppikirjat	✓	✓	✓
Vaihtoehtoiset oppimateriaalit	✓	✓	✓
Harjoituskirjat	✓	✓	✓
Oppimispelit	✓	✓	
Pienryhmäopetus	✓	✓	
Yksityisopetus	✓	✓	
Opetuspaketit		✓	✓
Oppimiskeskukset ja -laboratoriot		✓	✓
Tietokoneavusteinen opetus			✓

Korrektiivien opiskelun jälkeen opiskelijat voivat tehdä rinnakkaisen formatiivisen kokeen, jolla tarkistetaan oliko korrektiiveista hyötyä. Toisen kokeen tarkoituksena on myös antaa opiskelijoille toinen mahdollisuus ja toimia näin motivaattorina.

Ensimmäisessä kokeessa tavoitetasoon yltäneet voivat tehdä *syventäviä* aktiviteetteja. Näiden ei ole pakko suoraan liittyä oppimisosion asioihin.

Guskey (1985) esittää sopiviksi syventäviksi aktiviteeteiksi mm. vertaisopetusta, harjoitustehtävien tekoa toisille opiskelijoille, opeteltaviin taitoihin liittyvien media-aineistojen

tekoa, projekteja ja kirjoitelmia, haastavampia pelejä, ongelmia tai kilpailuja ja edistyneempää tietokoneavusteista opetusta.

Yksilöllisen opetuksen järjestelmä

Yksilöllisen opetuksen järjestelmää (engl. Personalised System of Instruction (PSI)), joka tunnetaan myös nimellä *Keller Plan* kehittäjänsä Fred. S. Kellerin mukaan, voidaan pitää yhtenä tavoiteoppimisen toteutuksena. Kuten LFM:ssä, opiskeltava sisältö on jaettu oppimisosioihin, joita seuraa formatiiviset kokeet. Seuraaviin oppimisosioihin saa edetä vasta, kun on osoittanut saavuttaneensa tavoitetason edellisissä. Erotuksena PSI:ssa opiskelu tapahtuu kuitenkin pääsääntöisesti omaan tahtiin ja itseopiskellen.

Opiskelumateriaali on tarjottu PSI:ssa perinteisesti kirjallisessa muodossa, jotta omaan tahtiin eteneminen olisi mahdollista. PSI kuitenkin kehitettiin 1960-luvulla, jonka jälkeen muunkinlaisen materiaalin kuten videoiden saavutettavuus on merkittävästi parantunut. PSI:in mukaisessa opetuksessa pidetään myös joskus luentoja, mutta niiden tarkoitus on enemminkin motivoida kuin olla ensisijainen oppimisen lähde.

Eräs PSI:in piirteistä on ollut *proktoreina* toimivat vanhemmat opiskelijat, jotka teetävät ja tarkistavat kokeet, ohjaavat ja nostattavat kurssin ilmapiiriä. Joissakin uudemmissa toteutuksissa proktorit on kuitenkin korvattu tietokoneavusteisella arvioinnilla. (Grant ja Spencer 2003)

1.2.3 Tavoiteoppimiseen liittyvä tutkimus

Tavoiteoppimisesta on ilmestynyt verrattain paljon tutkimuksia ja myös muutama meta-analyysi (esim. Guskey ja Gates 1986; C.-L. C. Kulik, J. A. Kulik ja Bangert-Drowns 1990).

279:sta eri analyysissä käsitellyistä tutkimuksista 224 (90%) raportoi vaikutusten olevan positiivisia oppimistulosten kannalta, efektiikoon vaihdellussa 0.24:n ja 0.94:n välillä. Lisäksi 51 (85%) 60:sta tutkimuksesta raportoi positiivisia vaikutuksia opiskelijoiden affektiivisiin piirteisiin (efektikoot 0.10-1.33). (S. A. Anderson 1994)

C.-L. C. Kulik, J. A. Kulik ja Bangert-Drowns (1990) analysoi 108 tutkimusta, joista 72:ssä tavoiteoppimista toteutettiin PSI- ja 36:ssa LFM-mallin mukaisesti. Kaikkiaan 103 tutkimusta 108:sta raportoi tuloksia kokeista, joista saatiin koetulosten keskiefektikooksi 0.52. 67 97:stä positiivisia vaikutuksia raportoineista tutkimuksista sisälsi tilastollisesti merkittäviä eroja, kun taas negatiiviset vaikutukset olivat tilastollisesti merkitsemättömiä. Kaiken kaikkiaan tulokset tukevat vahvasti olettamusta, että tavoiteoppimisella on positiivinen vaikutus oppimiseen.

Vaikutukset vaihtelevat kuitenkin luokkien sisällä vähemmän kyvykkäiden hyötyessä enemmän (0.60) verrattuna kyvykkäisiin opiskelijoihin (0.40). Tavoiteoppimisella on myös

todettu olevan teorian mukainen koetulosten hajontaa ja kyvykkyyden ja saavutusten välistä korrelaatiota vähentävä vaikutus.

Kaikki PSI-tutkimukset ja 19 LFM-tutkimuksista kohdistui korkeakouluopiskelijoihin. 19:stä korkeakoulutasoisesta LFM-tutkimuksesta 4 käsitteli matematiikan opetusta, joita vastaavat efektikoot olivat 0.28, 0.55, 0.57 ja 0.93 eli keskiarvoltaan 0.58. Samantyyppisiä tuloksia saatiin myös 11:stä matematiikan opetuksen PSI-tutkimuksista (efektikoko 0.02-0.96, keskiarvo 0.47).

Tavoiteoppiminen ja tietokoneavusteinen arviointi

Tietokoneavusteisen arvioinnin käyttöä tavoiteoppimisen kontekstissa esitettiin jo esim. Guskeyn 1985 kirjassa, ja joitakin tällaisia toteutuksia on myös raportoitu.

Boggs, M. Shore ja J. Shore (2004) toteutti tavoiteoppimista Blackboard-ympäristöä käyttäen matematiikan johdantokurssilla Marylandin Allegany-keuhkoulussa. Tällöin tunnistettiin neljä haastetta, joissa järjestelmä auttoi: useamman version tekeminen joka kokeesta, kokeiden pisteyttäminen, kokeiden aikatauluttaminen ja eri kohdissa oppimistavoitteita olevien opiskelijoiden opettaminen.

Useammat koeversiot luotiin arpomalla koekysymykset laajasta kysymyspankista, ja kokeiden pisteyttäminen onnistuttiin automatisoimaan tietokoneavusteisella arvioinnilla. Kokeiden aikataulutus ratkaistiin laittamalla kokeet tehtäväksi internetiin, ja halutessa salasanan taakse, jolloin kokeen voi tehdä valvotusti. Opiskelijat pystyivät näin etenemään omaan tahtiin käyttäen tavoitetasoa mittaavia kokeita ja verkosta löytyviä lisämateriaaleja.

Tavoiteoppimista tällä tavalla toteuttavalla kurssilla suurempi osa (66%, $N = 40$) läpäisi kurssin perinteiseen kurssiin verrattuna (55%, $N = 220$). Samankaltaisia tuloksia saatiin myös toisessa tutkimuksessa (Leonard, Hollos ja Gerace 2008), jossa tavoiteoppimisen verkkototeutus nosti koko vuoden kestävien piirianalyysikurssien läpäisyprosenttia 66%:sta 80%:iin vuoden takaiseen verrattuna.

Hoon, Chong ja Ngah (2010) tutki kolmea eri oppimismenetelmää: tietokoneavusteista tavoiteoppimista (computer-assisted mastery learning, CML), tietokoneavusteista vertais-tavoiteoppimista (computer-assisted cooperative mastery learning, CCML) ja tavoiteoppimiseen pohjautumatonta tietokoneavusteista vertaisoppimista (computer-assisted cooperative learning, CCL). Tutkimuksen tulokset viittasivat siihen, että tavoiteoppimiseen pohjautuvat menetelmät olivat tehokkaampia. Efektikoiksi saatiin 0.8778 (CCML verrattuna CCL:iin) ja 0.5603 (CML verrattuna CCL:iin). Vähemmän kyvykkäille opiskelijoille kuitenkin todettiin CML paremmaksi.

Capaldi (2014) toteutti omatahtista käänteistä tavoiteoppimista statiikan kurssilla käyttäen STEMSI-oppimisympäristöä. Luennot olivat saatavilla videoina ja luokkahuoneopetus oli varattu kokonaan harjoitustehtävien tekoa varten. Kurssi jaettiin kahteen-

toista oppimisosioon, ja opiskelijoiden piti osoittaa osaamisensa formatiivisilla kokeilla. Vaikka kurssilla opiskeltiin omatahtisesti, kurssia aikataulutti yhteiset väli- ja loppukokeet.

Oppimisosion opiskeluun käytetty aika vaihteli paljon opiskelijoiden välillä, ja hitaimman ja nopeimman opiskelijan ero kasvoi kurssin edetessä. Tämän takia mallin katsottiin sopivan parhaiten tilanteisiin, jossa kurssiaikataulussa on jonkin verran jouston varaa.

Em. tutkimuksissa ei kuvailtu arviointijärjestelmän ominaisuuksia tarkasti, mutta oletettavasti arviointijärjestelmät eivät olleet symbolisella laskentaohjelmistolla varusteltuja, vaan esim. monivalintatehtäviä tai numeerisesti tarkastettavia tehtäviä.

1.2.4 Kritiikki

Tavoiteoppiminen ja sen yhteys behaviorismiin

Behaviorismi on psykologian suuntaus, joka hylkää introspektiiviset tutkimusmenetelmät ja keskittyy tutkimaan käyttäytymistä, sen muuttamista ja sen havaittavissa olevia edeltäviä syitä ja seurauksia. (Phillips 2012) Behavioristisesta näkökulmasta oppimisen voi määritellä jonkin kokemuksen aiheuttamaksi muutokseksi käyttäytymisessä. (Burton, Moore ja Magliaro 1996) Monimutkaisempi oppiminen tapahtuu ketjutettujen käytösten oppimisen kautta. (Gagne 1985 kuten lainattu artikkelissa Burton, Moore ja Magliaro 1996)

Behavioristiset teoriat olivat vallitsevia teorioita kasvatustieteissä tavoiteoppimisen kehittämisen aikaan. Vaikka Bloom tai hänen kollegansa eivät koskaan suoraan puhunut tavoiteoppimisesta behavioristisena mallina, se kuitenkin usein nähdään perustuvan behaviorismiin (esim. Schunk 1996; Ertmer ja Newby 2013; Motamedi ja Sumrall 2000). Koska kasvatustieteellinen paradigma on myöhemmin suuntautunut vahvasti behaviorismista kohti kognitivismia ja konstruktivismia, tavoiteoppimisen näkeminen behavioristisena voi olla joillekin itsessään kritiikin aihe (esim. (Horton 1979)).

Tavoiteoppimisen behavioristiseksi koetut piirteet sisältävät välittömän palautteenannon, etenemisen estämisen ennen tavoitetasoon pääsyä, kriteeripohjaisen opiskelijoiden arvioinnin ja oppimisen jakamisen pieniksi, peräkkäisiksi askeleiksi. (Motamedi ja Sumrall 2000; Fosnot ja Perry 1996)

Fosnot ja Perry (1996) esittää, että tavoiteoppiminen "olettaa, että kokonaisuuden voi pilkkoa osiksi, että taidot voi pilkkoa osataidoksi, ja että nämä taidot voi asettaa 'oppimisjanelle'", jonka vuoksi tavoiteoppiminen olisi malliesimerkki behaviorismista.

Tavoiteoppimista puolustava Guskey (2008) hylkää osan tavoiteoppimista kohdistuvasta kritiikistä väärinymmärryksinä. Guskeyn mukaan jotkin varhaiset tavoiteoppimisen sovellusyritykset keskittyvät vain matalan tason kognitiivisiin kykyihin, pyrki hajottamaan oppimisen pieniksi kappaleiksi ja pitivät tiukasti kiinni tavoitevaatimuksista, ennen

kuin opiskelijoita päästettiin etenemään opiskelussa. Nämä kokeilut perustuivat "Bloomin teorian kapeaan ja virheelliseen ymmärrykseen", eikä "kuitenkaan Bloomin mistään kirjoituksista löydy tällaista kapea-alaisuutta ja jäykkyyttä".

Vaikka LFM ja PSI esittävät sisällön jakamista pienemmiksi oppimisosioiksi, "oppimisosiot tulisi ihanteellisesti määräytyä oppimateriaalin luonnollisten taukojen tai merkityksellisten sisältökokonaisuuksien mukaan". (Guskey 1985) Guskey myös jakaa huolen siitä, että "jos oppimisosiot tehdään liian lyhyiksi, oppiminen muuttuu katkonaiseksi ja yleistysten tekeminen vaikeaksi".

Se kuinka tiukasti tavoitevaatimuksista pidetään kiinni vaikuttaa riippuvan toteutuksesta. PSI tai omaan tahtiin etenevät toteutukset ovat usein tiukempia tässä suhteessa kuin LFM tai ryhmämuotoiset toteutukset. LFM:ssä formatiivista koetta voi yrittää kerran uudestaan, jonka jälkeen opiskelija siirtyy seuraavaan oppimisosioon tavoitteiden täyttymisestä riippumatta.

Tavoiteoppiminen ei ota kantaa siihen, miten opetus tulisi järjestää tai minkä tyyppisiä oppimistavoitteita kurssilla tulisi olla. Guskey huomauttaa tavoiteoppimisen olevan tutkimuksen mukaan olevan erityisen tehokas, kun sitä sovelletaan opetukseen, joka keskittyy korkeamman tason oppimistavoitteisiin kuten ongelmanratkaisuun, johtopäätösten tekemisiin, deduktiiviseen päättelyyn ja luovaan ilmaisuun. Tätä painotti myös Bloom itse.

Periaatteessa tämä tekisi mahdolliseksi soveltaa tavoiteoppimista minkälaiseen opetukseen ja oppimistavoitteihin tahansa, kunhan löytyy jokin tapa arvioida oppimista ja määritellä tavoitetaso. Siksi tavoiteoppiminen ei välttämättä ole luonnostaan epäyhteensopiva behaviorismin jälkeisten oppimisteorioiden kanssa, vaikka näitä harvoin yritetäänkään yhteensovittaa.

Toimiiko tavoiteoppiminen oikeasti ja ketä se hyödyttää?

Parhaan näytön synteesi peruskoulutuksen LFM-ohjelmiin keskittyvistä tutkimuksista Slavin (1987) päätyi "radikaalisti erilaisiin tuloksiin" kuin aiemmat meta-analyysit. Slavin painotti, että kun oppimistuloksia vertaillaan standardisoiduilla eikä tutkijan tekemillä kokeilla, tavoiteoppimis- ja kontrolliryhmät menestyvät samalla tavalla. Myös tutkijan tekemistä kokeista (17 tutkimusta) Slavin sai efektikoon keskiarvoksi 0.25, mikä on merkittävästi pienempi kuin esim. Guskeyn 1986 laskema yläasteella opiskelevia koskeva 0.93.

Toinen Slavinin huolenaihe oli, että joissakin tutkimuksissa koeryhmä sai merkittävästi enemmän opetusta kuin kontrolliryhmä. Tutkimukset, joissa erot annettavan opetuksen määrässä olivat äärimmäisiä, suljettiin synteessin ulkopuolelle. Slavin totesi kuitenkin, että näyttö tässä asiassa oli epäselvää, eikä löytänyt merkittäviä eroja efektikoossa saman verran ja eri verran opetusta antaneiden tutkimusten välillä.

Bloom (1987), Guskey (1987) ja L. W. Anderson ja Burns (1987) vastasivat Slavinin

kritiikkiin.

Bloom esitti, että tutkijoiden tekemät kokeet mittaavat tarkoitettuja oppimistavoitteita tarkemmin kuin standardoidut kokeet, mikä selittäisi miksi näin saadaan positiivisempia tuloksia. Bloomin mukaan pitkäaikaistutkimukset myös näyttäisivät tavoiteoppimisen parantavan standardoitujen kokeiden tuloksia.

Guskey kritisoi Slavinin tutkimusten valintakriteerejä ja piti "parhaan näytön synteesin" ajatusta "parhaasta" hyvin subjektiivisena, puolueellisena ja harhaanjohtavana.

Anderson ja Burns nostivat esille joitakin ongelmia efektikokojen laskutavassa. Yksi synteesin tutkimuksista sisälsi 18 efektikokoa -0.51:stä 0.54:ään, joista vain neljä oli välillä -0.20 ja 0.20. Näin saadun keskiarvon 0.04 käyttö tunnuslukuna oli heidän mielestään harhaanjohtavaa. Esille nostettiin myös kysymys efektikokojen soveltuvuudesta tavoiteoppimisen onnistumisen arvioinnissa. Yksi tavoiteoppimisen tavoitteista on pienentää oppimistulosten hajontaa, mitä efektikoot eivät mittaa.

Yksi Bloomin ja hänen kollegojensa väitteistä oli, että opiskelijoiden väliset erot oppimiseen vaaditussa ajassa pienentyisivät suotuisissa oppimisolosuhteissa. Arlin (1984) päätyi kuitenkin siihen, että vastakkaiselle väitteelle olisi uskottavampaa näyttöä. Tällöin tavoiteoppiminen saattaisi pidätellä nopeampia oppijoita, ja sisältäisi ajankäytön ja oppimistulosten välisen "dilemman".

Mueller (1976) esitti samanlaisia huolia. Mueller ehdotti, että *tavoitetasolle yltämisen nopeutta* pitäisi harkita myös yhtenä oppimistulosten mittarina. Mueller esitti myös, että tavoiteoppimisen nopeimmille oppijoille asettamasta katosta voisi päästä eroon jakamalla oppimistavoitteet kahteen tai useampaan kategoriaan, esim. perusteisiin ja edistyneempiin tavoitteisiin.

Lai ja Biggs (1994) tutki eroavaisuuksia oppimisessa ja affektiivisissa lopputulemissa pintasuuntautuneiden ja syväsuuntautuneiden oppijoiden välillä, kun käytössä oli tavoiteoppimisen malli. Vaikka LFM kasvatti koepisteiden keskiarvoa, se suosi pintaoppijoita. Ero myös kasvoi oppimisosiosta toiseen pintaoppijoiden pärjäävän paremmin ja syväoppijoiden huonommin. LFM oli myös affektiivisesti polarisoiva: pintasuuntautuneet oppijat yleensä ottaen pitivät mallista ja syväsuuntautuneet taas eivät.

Vaikka tutkimus ei sitä vahvistanut, vaikutti todennäköiseltä, että tavoiteoppiminen jossain määrin kannustaa pintasuuntautumiseen. Samalla kuitenkin mainittiin, että tavoiteoppimisen voisi saada tukemaan korkean tason ajattelua, jos formatiiviset kokeet tehtäisiin sopivimmiksi.

1.2.5 Nykyinen käyttö

Harvoja oppimismenetelmiä on toteutettu yhtä laajalti kuin tavoiteoppimiseen liittyviä. (Guskey 2008) Tavoiteoppimisen suosio on kuitenkin esim. Morganin (2011) mukaan tipunut, ja hän antaa tälle kolme mahdollista syytä:

1. Kouluttamattomat opettajat huonojen materiaalien kanssa. Kaikki kustantajien PSI:ia varten kehittämät materiaalit eivät olleet hyvin suunniteltuja, ja opettajat valittivat niiden mukana tulevien arviointityökalujen keskittyvän matalalle tasolle.
2. Tavoiteoppimisen toteuttaminen vaatii opettajalta paljon vaivaa varsinkin ensimmäisellä kertaa. Tämä voi olla este tavoiteoppimisen käyttöönotolle.
3. Tavoiteoppiminen on voinut toimia "liian hyvin". Morgan lainaa psykologian professori Howard Gallupin esimerkkiä:

Henry Pennypacker onnistui Floridan yliopistossa muuttamaan kahden ensimmäisen vuoden peruskurssit PSI-muotoon. Floridan sisäänottopoliitiikka salli tuolloin kenen tahansa lukion suorittaneen sisäänpääsyn yliopistoon. Ensimmäisen vuoden päätteeksi opinnot keskeytti tyypillisesti n. 80%. PSI-muotoisessa opetuksessa luku tippui 15%:iin. Tämän johdosta yliopistolla oli useampi tuhat opiskelijaa, jotka olivat halukkaita jatkamaan opintojaan edistyneemmille kursseille. PSI-muotoinen opetus lopetettiin kahden vuoden jälkeen.

Vaikka tavoiteoppimisen nimenomainen käyttö on ehkä vähentynyt, monet tavoiteoppimisen osista on yhä laajalti käytössä. Näistä Morgan nimeää suoritukseen perustuvan palautteen, tehokkaat oppimisperiaatteet, kognitiiviset tuet (engl. scaffolding) ja sisällön eriyttämisen esitietojen perusteella.

Tavoiteoppimista kohtaan on herännyt myös jonkin verran kiinnostusta uudestaan. Kuten kappaleen 1.2.3 katsauksessa huomattiin, tietokoneavusteinen arviointia on aiemmin käytetty tavoiteoppimisen työteliäisyyden keventämiseksi. Muita esimerkkejä on *käänteinen tavoiteoppiminen* (engl. flipped mastery), joka yhdistää tavoiteoppimisen käänteiseen oppimiseen (Bergmann ja Sams 2012). Simulaatiopohjainen tavoiteoppimista on viime aikoina käytetty lääketieteellisessä opetuksessa, ja tästä tehty tutkimus näyttää oppimistulosten parantuneen (McGaghie, S Barry Issenberg et al. 2011) ja myös joitakin lupaavia tuloksia hoidon tasolla (McGaghie, Saul B Issenberg et al. 2014).

Suomessa Bloomin työ tavoiteoppimisen parissa on toiminut inspiraation lähteenä Pekka Peuran kehittämässä *Yksilöllisen oppimisen menetelmässä*. (Toivanen 2013) Peuran menetelmässä korvataan formatiiviset kokeet harjoitustehtävien itsearvioinnilla ja opiskelu on omatahtista, minkä takia se muistuttaa enemmän PSI:ia kuin LFM:ä. Menetelmässä hyödynnetään myös vertaisoppimista ja sulautettua tai käänteistä oppimista verkosta löytyvien videoiden avulla.

1.3 Oppimisanalytiikka ja opetuksen tiedonlouhinta

Oppimisanalytiikka (engl. learning analytics) ja *opetuksen tiedonlouhinta* (engl. educational data mining) ovat läheisesti toisiinsa kytkeytyviä tutkimusaloja, jotka pyrkivät soveltamaan analytiikan ja tiedonlouhinnan menetelmiä oppimisen ja opetuksen kontekstissa. Oppimisanalytikkojen järjestö (The Society for Learning Analytics Research) määrittelee oppimisanalytiikan "oppijoita koskevan datan mittaamiseksi, keräämiseksi, analysoimiseksi ja raportoinniksi, tarkoituksena ymmärtää ja optimoida oppimista ja oppimisympäristöjä". Vastaavasti opetuksen tiedonlouhijoiden järjestö (The International Educational Data Mining Society) määrittelee opetuksen tiedonlouhinnan "tieteenalaksi, joka keskittyy kehittämään menetelmiä opetuksen yhä suuremman luokan datan tutkimiseen ja käyttämään näitä menetelmiä opiskelijoiden ja oppimisympäristöjen paremmin ymmärtämiseksi". (Siemens ja Ryan SJ d Baker 2012)

Monet menetelmät ovat käytössä molemmilla aloilla, mutta oppimisanalytiikka keskittyy enemmän ihmisen tekemään datan tulkintaan ja visualisointiin, kun taas opetuksen tiedonlouhinnassa keskitytään automatisoituihin menetelmiin. Käytettyjä menetelmiä paremmin erot tulevat esiin painotuksissa, tutkimuskysymyksissä ja menetelmien käyttöta-voissa. (Ryan Shaun Baker ja Inventado 2014)

Oppimiseen liittyvän datan lisääntymistä viime aikoina selittää paljon sähköisten oppimisympäristöjen käyttöönotto. Nämä keräävät tavanomaisten arviointikohteiden lisäksi myös paljon muuta dataa opiskelijan aktiivisuudesta oppimisympäristössä, kuten oppimateriaalin avaamista tai keskustelufoorumiviestien määrää. Oppimisympäristöalustoihin kuten Moodleen (2017) on myös myöhemmin ilmestynyt lokitietojen lisäksi valmiita työkaluja oppimisanalytiikkaa varten.

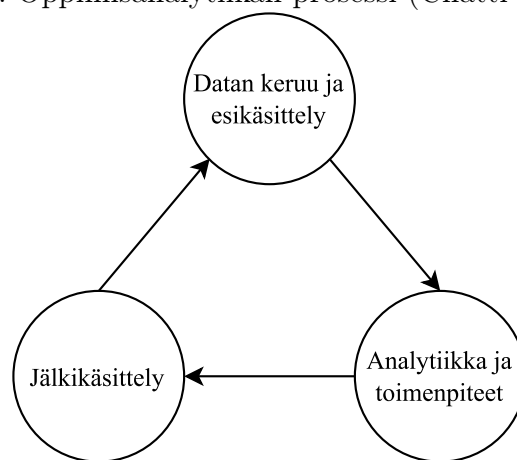
Eräs tärkeä menetelmien joukko on ennustava mallintaminen, jotka Ryan Shaun Baker ja Inventado (2014) jakaa kolmeen kategoriaan: luokitteleminen, regressio ja latentin eli piilevän osaamisen estimointi (engl. latent knowledge estimation). Luokittelevassa ennustamisessa pyritään ennustamaan aikaisempien havaintojen pohjalta mihin luokkaan uusi havainto kuuluu, kun taas regressiivisillä menetelmillä ennustetaan jatkuvia muuttujia. Latentin osaamisen estimoinnilla tarkoitetaan sellaisten mittareiden käyttämistä osaamisen mittaamiseen, jotka mittaavat osaamista epäsuorasti.

Tyypillinen sovelluskohde opetuksen kontekstissa on ennustaa esim. opiskelijan kursista läpikäymistä tai opintojen keskeyttämistä, jolloin opiskelijalle voidaan antaa oikea-aikaista tukea. Purduen yliopistossa kehitetty kurssilla menestymistä kolmiportaisesti ennustava Course Signals -järjestelmä lisäsi ylimpien arvosanojen määrää 10.37 prosenttiyksikköä ja vähensi koulutuksen keskeyttämistä. Arnold ja Pistilli (2012)

Chatti et al. (2012) jakaa oppimisanalytiikan prosessin kolmivaiheiseksi iteroitavaksi sykliseksi. Ensimmäisessä vaiheessa data kerätään ja esikäsitellään sellaiseen muotoon, että sitä on mahdollista käyttää valitun menetelmän syötteenä. Toisessa vaiheessa dataan so-

velletaan oppimisanalytiikan menetelmiä sellaisten piilevien rakenteiden löytämiseksi, jotka voivat hyödyttää oppimista. Analytiikkavaihe sisältää myös saadun tiedon perusteella tehdyt toimenpiteet, kuten seurannan, analyysin, ennustamisen, interventiot, arvioinnin, mukauttamisen, yksilöllistämisen, suositukset ja reflektion. Kolmannessa jälkikäsittelyn vaiheessa dataa voidaan yhdistää muihin lähteisiin ja määritellä seuraavaa iteraatiota varten tarvittavia muuttujia tai menetelmiä.

Kuva 1.9: Oppimisanalytiikan prosessi (Chatti et al. 2012)



Luku 2

Tutkimus

2.1 Tausta

Tässä tutkielman luvussa esitellään kirjoittajan Aalto-yliopistossa 2016-2017 tehdyn tavoiteoppimisen tutkimuksen (Pelkola, Rasila ja Christopher Sangwin 2018) tuloksia. Tutkimuksessa toteutettiin tavoiteoppimista tietokoneavusteista arviointia käyttäen.

Differentiaali- ja integraalilaskenta 1 (MS-A010X) on yhden muuttujan differentiaali- ja integraalilaskentaa ja differentiaaliyhtälöitä käsittelevä kuusiviikkoinen (5 op) kurssi Aalto-yliopistolla. Kurssi järjestetään pitkälle samansisältöisenä kaikissa tekniikanalan opinto-ohjelmissa.

Toteutukset MS-A0106 ja MS-A0107 valittiin Sulautetun matematiikan tavoiteoppimisen (Blended Mastery Learning in Mathematics) pilottikursseiksi. Pilotti oli osa verkkooppimisen kehittämisprojekti A!OLEa (Aalto Online Learning).

Kurssilla oli viikottaista opetusta neljä tuntia luentoina ja neljä tuntia laskuharjoitus-tilaisuuksina. Lisäksi kurssilla oli tietokoneavusteisesti arvioitavia verkkotehtäviä, paperille tehtäviä harjoitustehtäviä, oppimisosion päätteeksi tehtäviä tietokoneavusteisesti arvioituja formatiivisia kokeita ja kurssin lopuksi paperilla tehtävä kurssikoe. Kurssi jaettiin kahteen kolmeviikkoiseen oppimisosioon, joista ensimmäinen käsitteli raja-arvoja, sarjoja ja differentiaalilaskentaa, ja jälkimmäinen integraalilaskentaa ja differentiaaliyhtälöitä.

Bloomin LFM-mallia sovellettiin kurssin ydintaitojen oppimiseen tietokoneavusteisesti arvioituja verkkotehtäviä ja formatiivisia kokeita hyväksikäyttäen. Tämän seurauksena sovellettu malli poikkeisi joltain osin alkuperäisestä LFM-mallista.

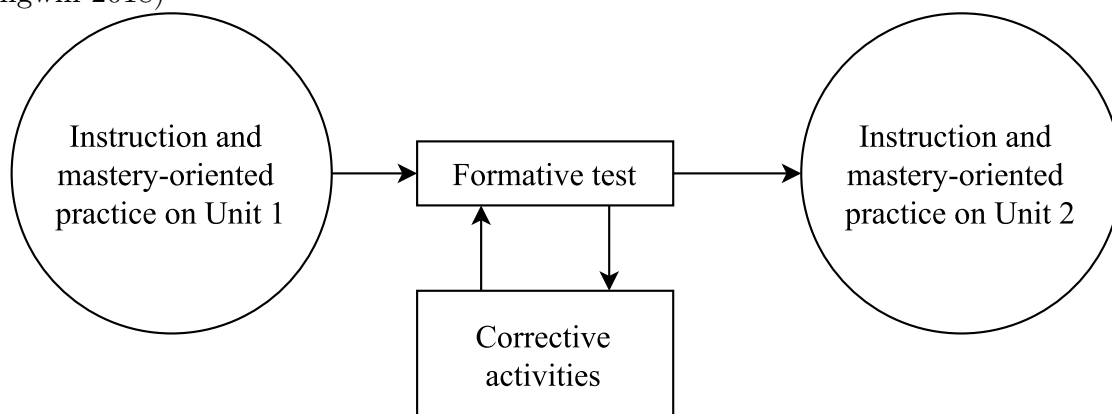
Alkuperäisessä LFM-mallissa tavoitteiden täyttymistä arvioidaan pelkästään formatiivisin kokein. Usein nämä kokeet ovat valvottuja monivalintakokeita, ja niissä sallitaan yksi uusimiskerta kokeen toisella versiolla.

Pilotoidussa mallissa tavoitteiden täyttymistä arvioitiin myös harjoitustehtäviä käyttäen. Formatiivisista kokeista oli vain yksi versio, uusimiskertoja ei ollut rajoitettu, eikä

oikein saatuihin koetehtäviin tarvinnut vastata toista kertaa. Formatiivisia kokeita kutsuttiin kurssilla tutummin "harjoituskokeiksi".

Oppimisosiot olivat hieman tavanomaista pidempiä, jotta formatiivisten kokeiden aiheuttama työtaakka saatiin pysymään kohtuullisena. Verkkotehtävät ja formatiiviset kokeet eivät myöskään sisältäneet kaikkia tehtävätyyppejä (esim. todistustehtäviä) eivätkä arvioineet joitakin korkeamman asteen oppimistavoitteita (esim. synteesiä), koska näiden tietokoneavusteinen tarkastaminen on haastavaa.

Kuva 2.1: Pilotoitu LFM-vaikutteinen tavoiteoppimismalli (Pelkola, Rasila ja Christopher Sangwin 2018)



Pilotin formatiivisia kokeita ja verkkoharjoituksia varten kehitettiin uusia STACK-kysymyksiä. Molemmilla kursseilla oli käytössä samat tehtävät. Tavoitetaso asetettiin 75-80%:iin tehtäväpisteistä, ja se laskettiin erikseen formatiivisille kokeille ja viikoittaisille verkkotehtäville.

Verkkotehtävät tai formatiiviset kokeet eivät olleet varsinaisesti pakollisia, mutta molemmat muodostivat osan kurssin loppuarvioinnista. Vain tavoitetason yltävät tehtäväpisteet uusintayritykset mukaan lukien laskettiin arvosanaan mukaan. Tavoitetasoon kannustavan pisteytyksen lopputuloksena pistemäärät eivät vaihdelleet paljoa, eivätkä nämä pisteet juurikaan määränneet arvosanaa. Tämän vuoksi verkkotehtävien ja kokeiden arviointia on perusteltua pitää ensisijaisesti formatiivisena.

Verkkotehtävillä ja formatiivisilla kokeilla oli kurssilla hieman erilainen rooli, minkä takia ne olivat myös asetettu toimimaan hieman eri tavoin. Vaikka molemmat antoivat palautetta oppimisen etenemisestä, verkkotehtäviä käytettiin alustavaa harjoittelua varten, kun taas formatiiviset kokeet varmistivat, että tavoitetaso harjoitetuissa taidoissa oli todella saavutettu ja säilytetty. Sekä verkkotehtävissä että formatiivisissa kokeissa kysymyksiin sai yrittää vastata uudelleen rajattomasti ilman pistevähennyksiä, mutta verkkotehtävissä palaute oli välitöntä, kun taas formatiivisissa kokeissa palautteen sai vasta

kokeen palauttamisen jälkeen.

Verkkotehtävät sulkeutuivat joka viikon sunnuntai, ja sisälsivät viisi tehtävää viikon luentoihin liittyen. Formatiivisen kokeen pystyi tekemään oppimisosion pääteeksi ennen seuraavan viikon luentoa tai kurssikoetta. Laskinten, oppimateriaalien ja muiden apuvälineiden käyttöä ei koetilanteessa kontrolloitu, mutta niiden pois jättäminen oli suositeltavaa. Ensimmäinen formatiivisista kokeista sisälsi neljä kysymystä ja jälkimmäinen viisi.

LFM-mallissa formatiivinen arviointi suositellaan yhdistettävän sopiviin korrektiiveihin. Pilotissa korrektiiveina toimivat ensisijaisesti STACK-järjestelmän antama palaute. Formatiivisen kokeen kysymyksiin liittyi myös kolmansien osapuolien tekemiä videoita, joita pääsi katselemaan ensimmäisen koeyrityskerran jälkeen.

Tavoitetasoon jo yltäneet opiskelijat eivät oletettavasti hyödy formatiivisista kokeista samalla tavalla. Koska verkkotehtävät tuottavat jo itsessään formatiivista palautetta, tutkimuksessa selvitettiin myös mahdollisuutta korvata formatiiviset kokeet verkkotehtävien tuottamalla datalla oppimisen edistymisestä.

Koska verkkotehtävät ja formatiiviset kokeet harjoittivat ja mittasivat lähinnä proseduraalista osaamista, kurssilla otettiin vastapainoksi käyttöön myös "ohjattua oivaltavaa oppimista"projektit tehtävien muodossa. Harmonista värähtelyä käsittelevään projektiin kuului neljä paperille tehtävää projektitehtävää ja esseemuotoinen yhteenveto.

2.2 Tutkimusmentelmät

Yksittäisten opiskelijakohtaisten STACK-tehtävien pisteet ja yrityskerrat vietiin Moodle-oppimisympäristöstä tarkoitusta varten kehitetyllä työkalulla. Tämän tuottama dataa analysoitiin R-laskentaohjelmiston avulla.

95 134:sta kurssille ilmoittautuneesta opiskelijasta kurssilla MS-A0106 ja 118 198:sta ilmoittautuneesta opiskelijasta kurssilla MS-A0107 olivat antaneet luvan käyttää heidän dataansa tutkimusta varten. Näistä opiskelijoista yhteensä 176 oli avannut kaikki ensimmäisen oppimisosion tehtävät ja formatiivisen kokeen vähintään kerran, mikä tutkimuksessa laskettiin tehtävän yritykseksi. Vastaavasti 168 oli tehnyt näin toisessa oppimisosiossa. Näiden suoritusten tuottamaa dataa käytettiin ennustavaan mallintamiseen.

Sekä STACKin tuottamaa dataa, että kurssipalautetta käytettiin arvioimaan STACKin soveltuvuutta tavoiteoppimisen toteuttamiseen. Verrokkiryhmän ja esitestin puuttumisen vuoksi tavoiteoppimisen vaikutusta oppimistuloksiin ei selvitetty tutkimuksessa.

Ennustavaa mallintamista kokeiltiin eri luokittelumenetelmillä ja datan esikäsittelyllä Caret R-ohjelmistopakettia apuna käyttäen. Luokittelumenetelmien toimivuutta arvioitiin kymmenkertaisella ristiinvalidoinnilla kolmella toistolla. Samanlaisia tuloksia saatiin usealla luokittelumenetelmällä. Näistä esitellään logistista regressiota ("glm"caretissa)

käyttämällä saadut tulokset.

2.3 Tutkimuskysymykset

TK1 Soveltuuko STACK tavoiteoppimisen toteuttamiseen?

TK2 Pystyykö tavoitetason täyttymistä ennustamaan verkkotehtävien tuottaman datan perusteella?

2.4 Tulokset

2.4.1 Kuvailevat tilastot

Tavoitetaso määriteltiin neljäksi viidestä (80%) tai kolmeksi neljästä (75%) pisteeksi viikoittaisissa verkkotehtävissä tai formatiivisissa kokeissa. *Alustavalla osaamisella* viitataan tässä osuuteen opiskelijoista, jotka ovat saavuttaneet tavoitetason jo ensimmäisellä yrittämällä, *lopullinen osaaminen* taas osuuteen opiskelijoista, jotka ovat saavuttaneet tavoitetason jollakin yrityskerralla. Näiden muutos on osuus opiskelijoista, jotka saavuttivat tavoitetason ensimmäisen yrityksen jälkeen.

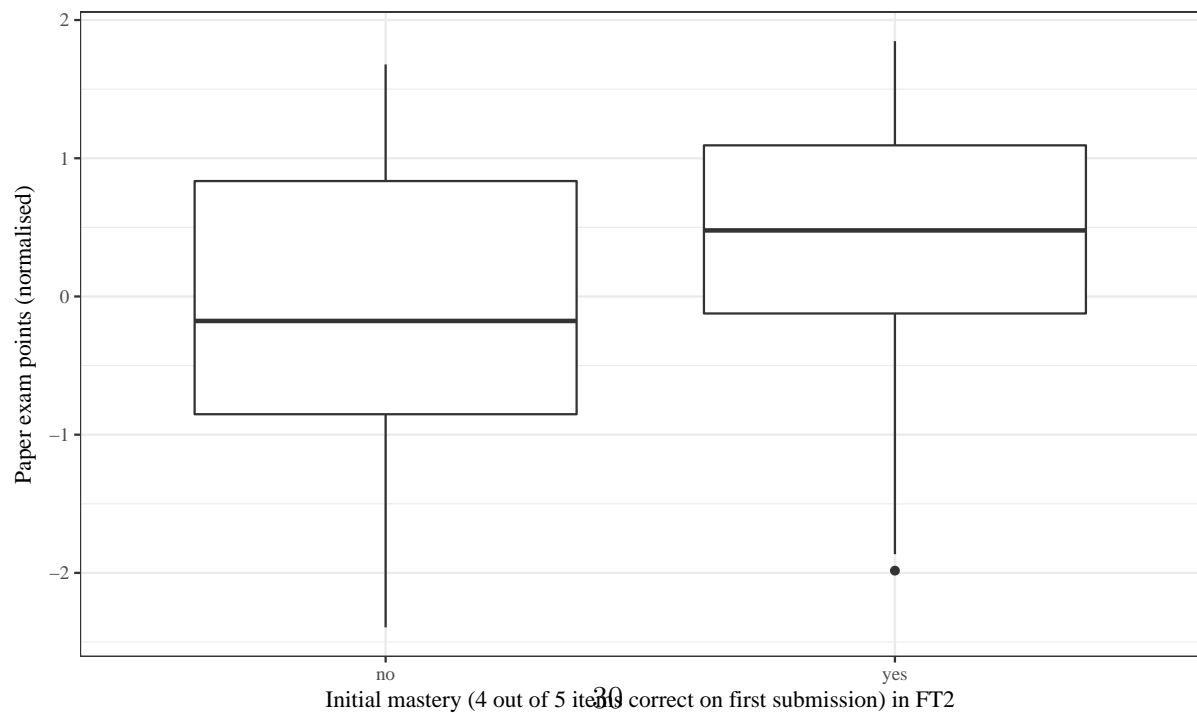
Tavoitetasotilastot esitellään taulukossa 2.1. Keskimäärin 88% verkkotehtäviä tehneistä ja 94% formatiiviseen kokeeseen osallistuneista saavuttivat lopulta tavoitetason. Ensimmäisellä yrittämällä tavoitetasoon pääsi 12% verkkotehtävissä ja 45% formatiivisissa kokeissa. Kuten voisi olettaa, alustava taso oli korkeampi formatiivisissa kokeissa kuin verkkotehtävissä, joita käytettiin aikaiseen harjoitteluun. Formattiivisten kokeiden alustava taso oli kuitenkin huomattavasti matalampi mitä verkkotehtävien lopullinen taso antaisi ymmärtää.

Kun opiskelijat jaettiin toisen formatiivisen kokeen tavoitetason täyttymisen mukaan, kurssikokeen pisteiden keskiarvoissa löytyi 0.51:n keskihajonnan kokoinen ero. (Taulukko 2.2) Tämä edelleen viittaisi siihen, että lopullinen osaaminen ei ole täysin verrattavissa alustavaan osaamiseen.

Taulukko 2.1: Prosenttiosuudet tavoitetasoon yltäneistä opiskelijoista

<u>Alustava osaaminen</u>	VT1	VT2	VT3	FK1	VT4	VT5	VT6	FK2
MS-A0106	13%	22%	7%	47%	6%	11%	8%	52%
MS-A0107	22%	25%	10%	44%	7%	13%	3%	38%
Yhteensä	18%	24%	8%	45%	6%	12%	6%	44%
<u>Lopullinen osaaminen</u>								
MS-A0106	92%	95%	85%	95%	79%	90%	90%	98%
MS-A0107	90%	90%	83%	90%	85%	90%	87%	94%
Yhteensä	91%	92%	84%	92%	82%	90%	89%	96%
<u>Muutos</u>								
MS-A0106	80%	73%	78%	47%	73%	78%	82%	46%
MS-A0107	67%	65%	74%	47%	78%	77%	84%	56%
Yhteensä	73%	69%	76%	47%	76%	77%	83%	52%

Kuva 2.2: Koepisteiden vertailua tavoitteeseen yltäneiden ja yltämättömien opiskelijoiden välillä (Pelkola, Rasila ja Christopher Sangwin 2018)



2.4.2 Kurssipalautekysely

MS-A0106:n kurssipalautekysely sisälsi neljä kysymystä likert-asteikolla kurssilla käytetystä tavoiteoppimisen mallista. (Taulukko 2.2) Palaute oli suurimmaksi osaksi mallin kannalta myönteistä.

Tavoitetasoon kannustava pisteytysmalli (vain tavoitetason ylittävien pisteiden laskeminen) vaikutti lisäävän (40% vastaajista) enemmän kuin vähentävän (14%) harjoittelua. 37% piti formatiivisia kokeita hyvin hyödyllisinä, kun taas vain 5% piti niitä täysin hyödyttöminä.

Formatiivisissa kokeissa korrektiiveina toimineita videoita piti hyödyllisenä 91% videoita katselleista (26% kaikista vastanneista). On epäselvää, miksi niin monet jättivät katsomatta videoita, mutta osuutta pitää verrata joka tapauksessa alustavaan osaamiseen (keskimäärin 50% MS-A0106:lla), koska videot olivat katseltavissa vasta ensimmäisen yrityksen jälkeen.

Koska formatiiviset kokeet eivät olleet valvottuja mutta kuitenkin tarkoitettu ratkaistavaksi ilman apuvälineitä tai oppimateriaaleja, "koevilppi" oli mallin kannalta huolenaihe. Suurin osa (84%) myönsi käyttäneensä apuvälineitä ja oppimateriaaleja formatiivisen kokeen aikana ainakin muutaman kerran, vaikkakin vain 13% sanoi tehneensä sitä usein. Virheellisistä vastauksista jopa triviaalisti laskimella ratkaistaviin tehtäviin voi päätellä, että ainakin ensimmäinen yrityskerta oli yleensä melko rehellinen.

Taulukko 2.2: Tavoiteoppimisen malliin liittyvät kysymykset kurssikyselyssä (89 vastaajaa)

	1	2	3	M	SD
Harjoituskokeista oli hyötyä 1) ei lainkaan 2) jonkin verran 3) paljon	5%	58%	37%	2.3	0.6
Oliko harjoituskokeeseen liittyvistä videoista hyötyä?* 1) ei lainkaan 2) jonkin verran 3) paljon	9%	41%	50%	2.4	0.7
Käytitkö harjoituskokeessa apuvälineitä (laskinta, kirjaa yms.) 1) en kertaakaan 2) pari kertaa 3) usein	15%	71%	13%	2.0	0.5
Tavoitetasoon kannustava pisteytystapa (0 pistettä jos < 80% tehtäväpaketista tehty) vaikutti tehtävientekooni 1) enimmäkseen negatiivisesti 2) ei mitenkään 3) enimmäkseen positiivisesti	14%	47%	40%	2.3	0.7
* sisältää vain vastaajat, jotka ilmoittivat katsoneensa videoita (26% formatiiviseen kokeeseen osallistuneista)					

Joitakin osia tavoiteoppimisen mallista myös kommentoitiin positiiviseen sävyyn kysymyksen "Mikä kurssilla oli hyvää? Mikä edisti oppimistasi?" vastauksissa.

Lähes kaikki 74:stä vastauksesta mainitsivat laskuharjoitukset tai laskuharjoitustilaisuudet jossain muodossa. 18 mainitsi STACK-verkkoharjoitukset erityisesti ja 8 formatiiviset kokeet. Alle on kerätty esimerkkejä vastauksista.

Stack-tehtävät ja harjoituskokeet olivat hyvä lisä, ylipäättään kaikki ylimääräinen kotiintuotava auttaa, sillä omalla kohdalla perusasioiden jankkaamista pitäisi painottaa tehtävissä hieman enemmän ennen itse soveltavia tehtäviä.

Välikokeista sai hyvän käsityksen miten hallitsee kurssin asiat

Harjoituskokeet pakottivat kertaamaan

Tavoitetasoon kannustava pisteytystapa sai myös maininnan:

kurssilla oli hyvää se, että stack tehtävien tekeminen oli ns pakollista

Edelliselle kysymykselle oli myös vastine "Mikä kurssilla oli huonoa/ei toiminut? Mikä haittasi oppimistasi?". 68:sta vastauksesta suurin osa keskittyi projektitehtäviin, luentoihin ja luentokalvoihin. Kahden vastaajan mukaan kurssilla oli liian monen tyyppisiä tehtäviä.

STACK-tehtäviä pidettiin sekä liian vaikeina että liian helppoina.

... osa stackeistä oli sellaisia, että en kyllä itse nähnyt niissä perustehtävän häivääkään, ainakaan luentojen perusteella ei ollut mitään käsitystä ratkaisumalleista eikä assarien selitystenkään jälkeen aina ollut ymmärtänyt

Laskaritehtäviä oli ihan liikaa ja kaikki oli haasteeltaan heikkoja. Mieluummin vaikka 3x vähemmän, mutta haastavempia. Etenkin stack-tehtävät tuntuivat usein ajanhukalta.

Formatiivisia kokeita ei kritisoitu ohjeistuksen epäselvyyttä lukuun ottamatta.

2.4.3 Ennustemallit

Tavoitetasen täyttymisen ennustaminen formatiivisissa kokeissa verkkotehtävien perusteella osoittautui odotettua haastavammaksi.

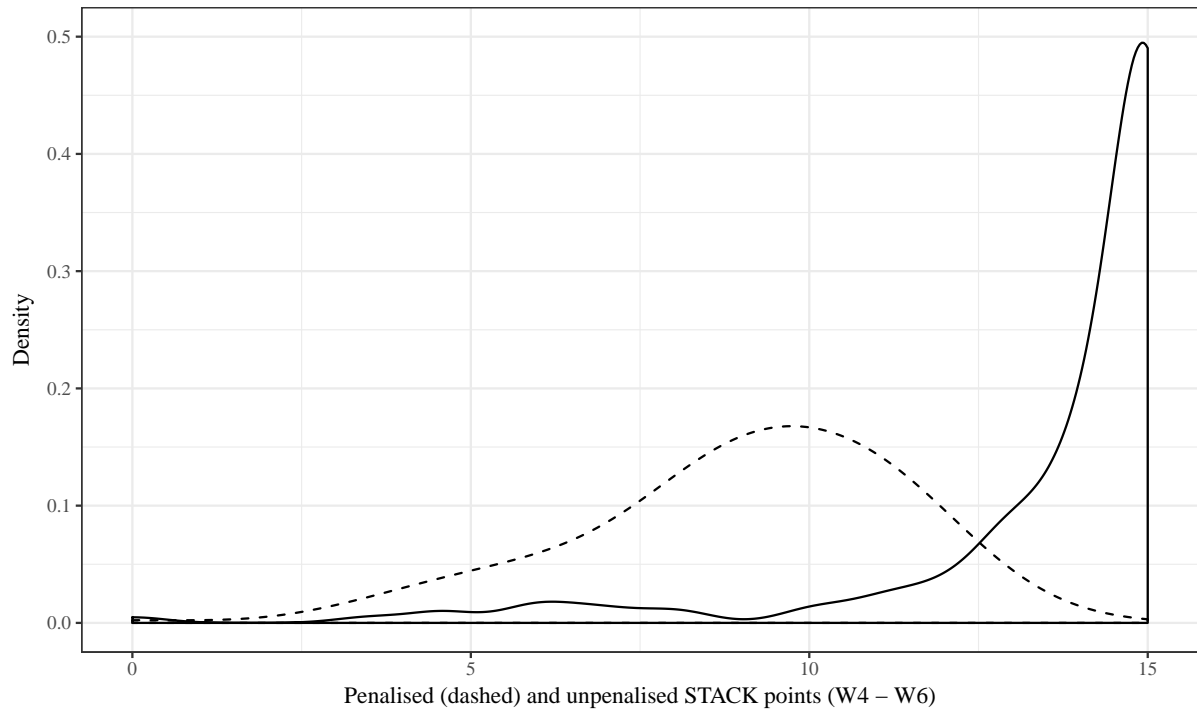
Ilman pistevähennyksiä uudelleenyrityksistä tehtäväpisteasteikko loppui kesken asteikon yläpäässä. Pistevähennyksiä simuloitiin jälkeenpäin kaavalla:

$$(2.1) \quad \text{vähennetyt pisteet} = \text{floor}(\text{tehtäväpisteet}) \cdot 0.7^{\text{uudelleenyrityskerrat}},$$

jossa $\text{floor}(x)$ pyöristää osapisteet nolaksi. Kaavalla lasketut pisteet muodostivat vähemmän vinon jakauman. (Kuva 2.3) Verkkotehtävien 4-6 vähennetyillä pisteillä oli myös korkeampi Spearman-korrelaatio (0.51) kurssikokeen pisteiden kanssa verrattuna vähentämättömiin pisteisiin (0.40).

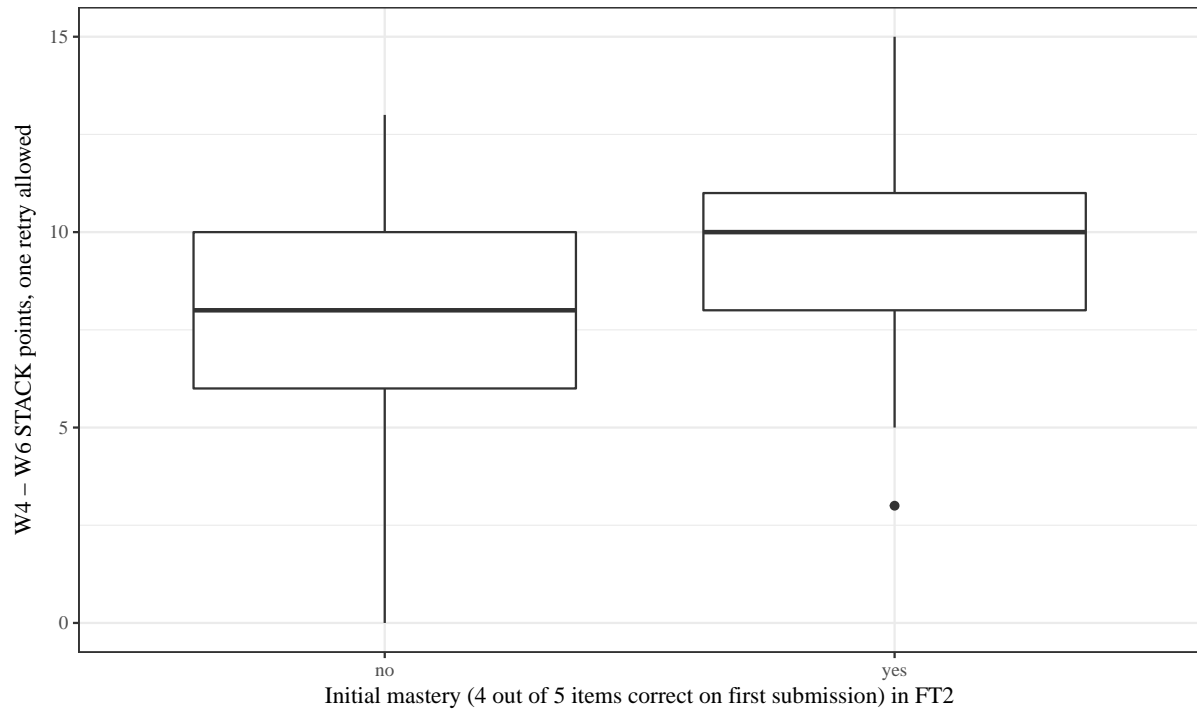
On syytä huomata, että todelliset pistevähennykset tai vastauskertojen rajoittaminen saattaisi vaikuttaa opiskelijan käyttäytymiseen, eikä simuloitujen pistevähennykset vastaa välttämättä täysin tällaista tilannetta. Aineistossa havaittiin korkeita yrityskertojen määriä, mikä saattaa viitata siihen, että osa opiskelijoista yrittää ratkaista tehtäviä kokeilemalla, kun siitä ei rangaista.

Kuva 2.3: Pisteasteikon loppuminen viikoilla 4-6 (Pelkola, Rasila ja Christopher Sangwin 2018)

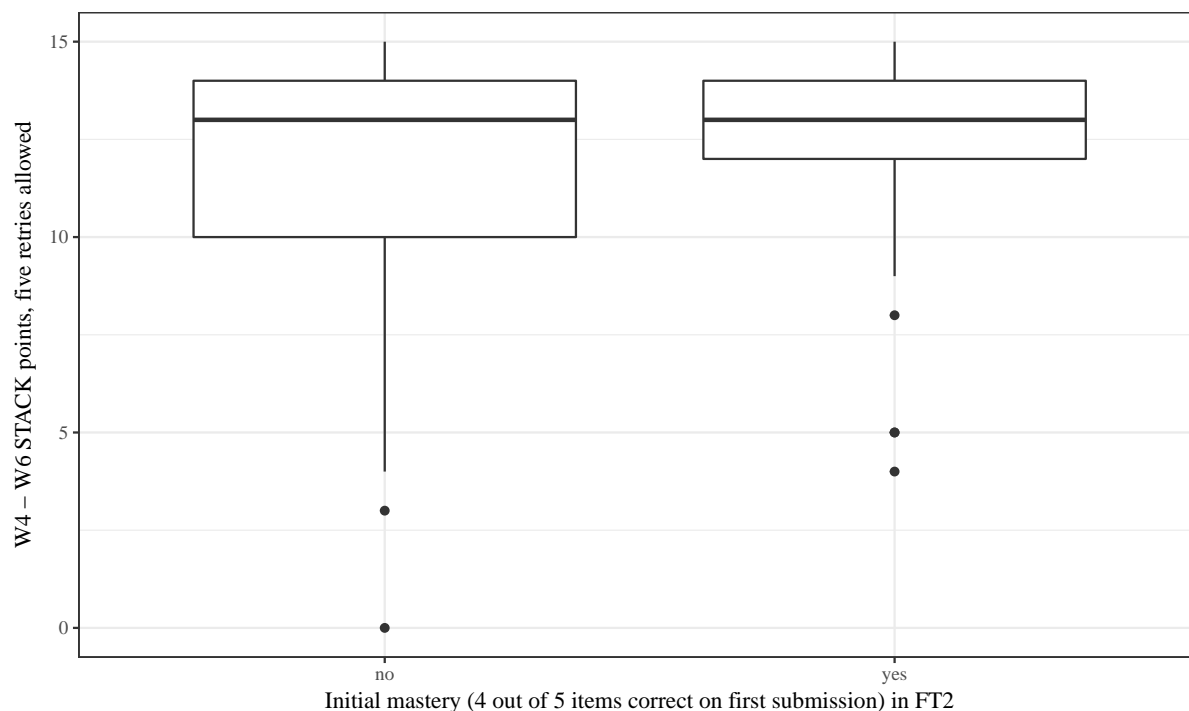


Verkkotehtäväpisteitä ja tavoitetason täyttymistä formatiivisissa kokeissa vertailemalla nähdään, että tehtäväpisteet erottelevat osaamistasoa huonosti. Ero ryhmien välillä on suurimmillaan, kun tehtäväpisteisiin otetaan huomioon enintään yksi uusintakerta, mutta useammilla uusintakerroilla asteikon loppuminen tulee taas näkyviin.

Kuva 2.4: Kun yksi uusintakerta sallittiin pistelaskussa, tehtäväpisteet erottelivat opiskelijoiden tasoa jonkin verran (Pelkola, Rasila ja Christopher Sangwin 2018)



Kuva 2.5: Viiden uusintakerran jälkeen mediaanitehtävapisteeissä ei enää näy eroa (Pelkola, Rasila ja Christopher Sangwin 2018)



Kuten taulukosta 2.1 voitiin päätellä, lopullinen osaamistaso verkkotehtävissä ei muuttunut alustavaksi osaamistasoksi formatiivisissa kokeissa, ja pistemäärä ei myöskään vaikuttanut erottelavan tavoitetason toteutumista (kuvat 2.4 ja 2.5). Siksi formatiivisessa kokeessa tavoitetason täyttymisen ennustamista varten tarvittaisiin hienostuneempi malli. Tutkimuksessa pyrittiin luomaan tavoiteosaamisen täyttymisen luokittelijaa, jolla pystyttäisiin lopulta myös korvaamaan formatiiviset kokeet.

Ennustavaa mallintamista kokeiltiin useilla Caret-pakkauksesta löytyneillä menetelmillä. Logistinen regressio suoriutui luokittelusta yhtä hyvin kuin gradienttiteuttujen puiden (engl. gradient-boosted trees) kaltaiset kehittyneemmät menetelmät. Hyvänä puolena logistinen regressio myös kertoo myös luokkatodennäköisyydet, jotka mahdollistavat helpon luokittelurajan optimoinnin. Tässä tapauksessa väärän positiivisen tuloksen kustannusta (vaillinainen oppiminen) voidaan pitää suurempana kuin väärän positiivisen tuloksen kustannusta (turhasta kertaamisesta aiheutuva ajanhukka).

Viikkojen 4-6 verkkotehtävistä saatavaa dataa käytettiin alustavan osaamistasen ennustamiseen. Vähennettyjen tehtävapisteen (kaava 2.1) summa tuotti lopulta parhaan lopputuloksen.

On huomioitava, että ylisovittamisen takia täysien havaintojen lukumäärä (168 toisessa oppimisosiossa) rajoittaa selittävien muuttujien lukumäärää. Datan vähäinen määrä saattaa selittää myös sitä, miksi yksittäiset tehtäväpisteet ja yrityskerrat eivät tarkentaneet mallia.

Datan harkittua esikäsittelyä tarvittiin myös jonkin verran, koska onnistumista ja luovuttamista edeltävien yrityskertojen lukumäärä mittaa olennaisesti eri asiaa. Kokeilun jälkeen päädyttiin eksponentiaaliseen pistevähennysmalliin (kaava 2.1), koska tämä mahdollisti uudelleenyritysten ja pistemäärien yhdistämisen samaksi muuttujaksi, eikä kärsinyt asteikon loppumisesta kesken samalla tavalla kuin lineaarinen pistevähennysmalli.

Näin saatu malli ennusti 0.64:n tarkkuudella, että opiskelija jäisi tavoitetasosta ensimmäisellä yrittämällä toisessa formatiivisessa kokeessa. Tämä on pieni parannus triviaaliin ennustemalliin verrattuna, jossa kaikkien opiskelijoiden oletettaisiin jäävän tavoitetasosta (0.56).

Taulukko 2.3: Luokittelijan sekaannusmatriisi (10-kertainen ristiinvalidointi, 3 toistoa)

<u>Oikea</u>		
<u>Ennuste</u>	alle tavoitetason	tavoitetaso
alle tavoitetason	33.7	17.1
tavoitetaso	18.7	30.6

2.5 Johtopäätökset

Soveltuuko STACK tavoiteoppimisen toteuttamiseen?

STACK-järjestelmä pystyy arvioiminaan suurimman osan kurssin Differentiaali- ja integraalilaskenta 1 oppimistavoitteista, ja on siis teoriassa sopiva väline tavoiteoppimisen toteuttamiselle tällä kurssilla. Erityisesti STACK mahdollistaa monien sellaisten tavoitteiden automaattisen arvioinnin, joita ei pystyisi arvioimaan luotettavasti monivalintakysymyksillä.

Toteutus on myös osoittautunut käytännössä toimivaksi siinä mielessä, että jokaisessa formatiivisessa kokeessa ja verkkotehtäväpaketissa huomattava osa opiskelijoista ylsi tavoitetasoon, vaikka ei ensimmäisellä yrittämällä siinä onnistunutkaan (69-83% verkkotehtävissä ja 47-52% formatiivisissa kokeissa).

Kurssipalautteen perusteella malli myös sai opiskelijoiden hyväksynnän. Formatiivisia kokeita pidettiin hyödyllisinä ja tavoitetasoon kannustava pisteytysmalli kannusti opiskelijoita tarkoitettulla tavalla.

Jonkin verran huolta aiheuttaa kuitenkin se, ettei verkkotehtävissä tavoitetasoon lopulta yltäminen johtanut tavoitetasoon täyttymiseen ensimmäisellä kerralla formatiivisissa kokeissa. Ensimmäisellä kerralla tavoitetasoon yltäneet myös pärjäsivät kurssikokeessa paremmin. Tätä saattaa selittää osiltaan ero *harjoitus-* ja *koeosaamisen* välillä.

Harjoitustehtävän ratkaiseminen voi olla monista syistä johtuen huomattavasti helpompaa kuin vastaavan tehtävän ratkominen koeolosuhteissa. Opiskelija saattaa saada harjoittellessaan apua toiselta opiskelijalta tai opettajalta, ei ole pelkästään muistamiensa asioiden varassa, voi tarkistaa vastauksensa ja yrittää uudelleen, sekä käyttää todennäköisemmin laskinten kaltaisia apuvälineitä.

Samaan tapaan saman kokeen uusintakerrat saattavat poiketa oleellisesti ensimmäisestä kerrasta.

Ei kuitenkaan ole selvää vastausta siihen, kumpi näistä osaamisen muodoista on varsinaisesti toivottua osaamista. Formatiiviset kokeet kuitenkin vaikuttavat paljastavan osaamisesta jotain, mitä pelkät harjoitustehtävät eivät näytä. Tämän vuoksi formatiivisia kokeita voi joka tapauksessa pitää oppimisille hyödyllisinä.

Alustavan ja lopullisen tavoiteosaamisen eroa voi pitää myös osoituksena korrektiivien tehottomuudesta tai siitä, että lopullinen tavoiteosaaminen on saavutettu apuvälineiden avulla.

Pystyykö tavoitetason täyttymistä ennustamaan verkkotehtävien tuottaman datan perusteella?

Opiskelijan suoriutuminen verkkotehtävissä ennusti formatiivisten kokeiden mukaista tavoiteosaamista pienissä määrin, ja tässä tapauksessa se ei oikeuta formatiivisten kokeiden korvaamista ennustemallilla.

Tulokset olivat kuitenkin positiivisia, ja on mahdollista, että ne paranisivat havaintojen lukumäärää kasvattamalla, toisilla selittäväillä muuttujilla tai mallin hienosäädöllä.

Jotkin aiemmin mainittuja huolenaiheet pätevät myös tässä. Formatiivisten kokeiden valvonta saattaisi tehdä koneoppimismallin opetukseen käytetystä datajoukosta luotettavamman.

Luku 3

Pohdintaa

3.1 Bloomin tavoiteoppimisen kehittäminen

Vaikka tavoiteoppiminen on ehkä poistunut muodista, sen päämäärät ja formatiivisen arvioinnin ydin ovat yhä ajankohtaisia. Tämän tutkielman tarkoitus oli selvittää, onko opetusteknologian kehittyminen tehnyt tavoiteoppimisen toteuttamisesta helpompaa. Se ei kuitenkaan anna suoraa vastausta siihen *pitäisikö* tavoiteoppimista ylipäättään käyttää tänä päivänä opetuksessa.

Tavoiteoppimisen toteutettavuus ja tunnetut hyödyt eivät ole välttämättä vielä riittäviä syitä käyttöönotolle, jos tavoiteoppimista pidetään itsessään liian rajoittavana tai jopa epäyhteensopivana nykyisten oppimiskäsitysten ja kouluopetuksen päämäärien kanssa.

3.1.1 Tavoiteoppimismallien oletusten väljentäminen

Tavoiteoppimisen taustamalleihin voi ajatella kuuluvan joitakin implisiittisiä ja mahdollisesti ongelmallisia oletuksia, kuten:

1. Osaamista ei pysty helposti menettämään, esim. unohtamalla tai väärinymmärrysten kautta
2. Formattiivisten kokeiden mukainen osaaminen on riittävä ja merkityksellinen oppimisen mittari
3. Kurssin oppimistavoitteet voidaan jakaa peräkkäisiin osioihin ja *osien hallitseminen* johtaa *kokonaisuuden hallintaan*
4. Oppimistavoitteet ovat hyvinmääriteltyjä ja samat kaikille opiskelijoille

Näiden oletusten kannalta ongelmallisia tilanteita saattaa olla mahdollista ratkaista muutoksilla käytettyyn tavoiteoppimisen malliin. Samalla on kuitenkin hyvä muistaa, että yhdestä tavoiteoppimisen mallista saadut tulokset eivät välttämättä ole suoraan yleistettävissä toisiin, joita täytyy mahdollisesti arvioida erikseen.

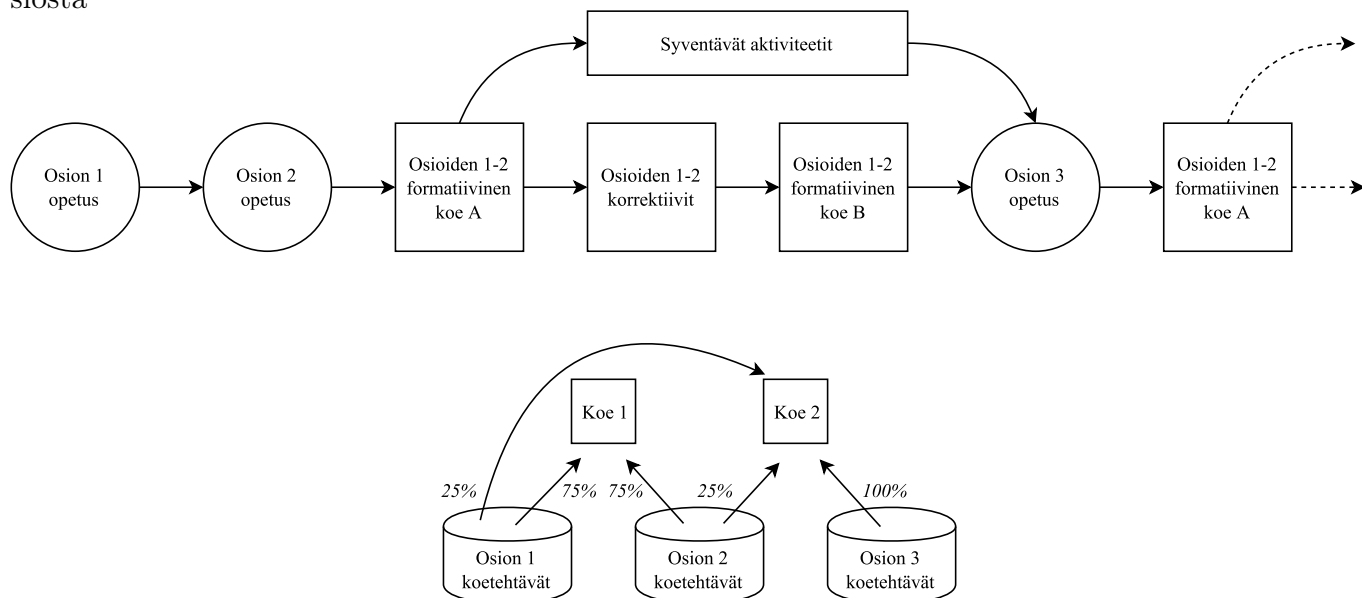
Harvennettu tavoiteoppiminen

Aikavälikertaus (engl. spaced repetition) on käytäntö, jossa aikaisemmin opittua kerrataan kasvavin välein. Tämän muistamista tukevaa ominaisuutta kutsutaan *harvennusefekti*ksi (engl. spacing effect), josta saattaisi olla hyötyä myös tavoiteoppimisen yhteydessä.

Rohrer ja Taylor (2007) sai tulokseksi, että yhden laskuharjoitustilaisuuden harjoitustehtävien hajauttaminen pitkin kurssia paransi oppimistuloksia. Tätä voisi soveltaa myös tavoiteoppimisen mallissa, jossa jokainen oppimisosion koe sisältää myös koekysymyksiä edellisistä osioista. Tavoitetaso tässä mallissa voitaisiin määritellä joko kokeen mukaan tai eritellä nykyisen ja edellisten osioiden osaamiseen. Tavoitetason täyttyminen voisi määräytyä myös viiveellä useamman kokeen aikana.

Kokeen mukaan määräytyvä osaamistaso olisi helpoin toteuttaa, mutta ei enää aivan sama asia kuin yksittäisen eheän osion hallinta. Se saattaisi myös epäonnistua erityisen vaativan oppimisosion osaamattomuuden tunnistamisessa. Koetuloksen jakaminen oppimisosioden mukaan ehkä taas vaatisi useampia koekysymyksiä, tai saattaisi heikentää mittarin tarkkuutta. Jos osaaminen määritellään viiveellä, korrektiivien ajoittaminen voi muodostua hankalaksi.

Kuva 3.1: Harvennettu tavoiteoppiminen ja koetehtävien jakaminen kolmesta oppimisosiosta



Vaihtoehtoiset arviointimenetelmät

Tavoiteoppimisen malleissa arvioidaan osaamista yleensä kokein, mutta tavoiteoppiminen mahdollistaa sinänsä muunkinlaisten arviointimenetelmien käytön. Arviointimenetelmän täytyy kuitenkin pystyä mittaamaan tai ennustamaan opiskelijan osaamista oppimisosion lopussa ja määrittämään jollakin tavalla osaamisen tavoitetaso. Arviointimenetelmän reliabiliteetti ja validiteetti ovat tavoiteoppimisen onnistumisen kannalta oleellisia asioita, joten näihin on myös syytä kiinnittää huomiota.

Yksi tapa arvioida osaamista on pitää kirjaa osion aikana tehdyistä harjoitustehtävistä, mutta kuten tutkimuksessa havaittiin, harjoitustehtävät antavat helposti osaamisesta koetta positiivisemmän kuvan. Tarkemman kuvan saamiseksi pitäisi huomioida jollakin tavalla kuinka haastava tehtävä opiskelijalle oli esim. yrityskertojen lukumäärän tai itsearvioinnin avulla.

Jatkuvan arvioinnin ongelma on myös, että osaaminen saattaa arvioinnin aikana muuttua, eikä sitä pysty aina ennustamaan. Tämän ratkaisemiseksi oppimisosion lopussa voi antaa myös erikseen arvioitavia kertaustehtäviä tai laajempia töitä opiskelijoiden tehtäväksi.

Osaamista voi oppimisosion lopuksi arvioida myös itsearvioinnilla kuten esim. Peuran yksilöllisen oppimisen menetelmässä tehdään. Yksi tapa toteuttaa tällainen itsearviointi

on antaa koetta vastaavat tehtävänannot, mutta ratkaisun sijaan opiskelijaa pyydetään arvioimaan osaako hän ratkaista tehtävät. Itsearviointin reliabiliteetti ja validiteetti suhteessa opettajan suorittamaan arviointiin voi kuitenkin vaihdella ja riippuu myös opiskelijasta. Itsearviointiin perustuvalla tavoiteoppimisella ei välttämättä ole myöskään täysin samanlaisia myönteisiä affektiivisiä vaikutuksia kuin formatiivisiin kokeisiin perustuvalla, koska ulkopuolinen “objektiivisemmaksi” koettu vahvistus osaamisesta jää puuttumaan.

Rinnakaiset oppimisosiot

Kurssin oppimistavoitteiden jakaminen peräkkäisiin osioihin ei onnistu aina luontevalla tavalla. Esim. lukion opetussuunnitelmassa mainitaan kurssin Polynomifunktiot tavoitteiksi, että opiskelija:

1. harjaantuu käsittelemään polynomifunktioita
2. osaa ratkaista toisen asteen polynomiyhtälöitä ja tutkia ratkaisujen lukumäärää
3. osaa ratkaista korkeamman asteen polynomiyhtälöitä, jotka voidaan ratkaista ilman polynomien jakolaskua
4. osaa ratkaista yksinkertaisia polynomiepäyhtälöitä
5. osaa käyttää teknisiä apuvälineitä polynomifunktion tutkimisessa ja polynomiyhtälöihin ja polynomiepäyhtälöihin sekä polynomifunktioihin liittyvien sovellusongelmien ratkaisussa

Tavoitteet 1-4 on luontevaa jakaa peräkkäisiin oppimisosioihin, kun taas tavoitetta 5 on luontevinta lähestyä pitkin kurssia. Tavoitteen 5 voisi ajatella muodostavan oman rinnakkaisen oppimisosionsa, jolloin teknisten apuvälineiden osaaminen voidaan varmistaa erikseen, kun muiden oppimisosioden lomassa arvioituna se saattaisi jäädä huomiotta.

Jossain mielessä rinnakaiset oppimisosiot muistuttavat yksinkertaisesti useampaa päällekkäistä kurssia. Liian monesta rinnakkaisesta oppimisosiosta saattaakin olla samalla tavalla oppimiselle haittaa kuin liian monesta päällekkäisestä kurssista, minkä vuoksi oppimisosioden rinnakkaisuutta pitäisi käyttää harkiten. Peräkkäiset oppimisosiot ovat perusteluja varsinkin silloin, kun edellinen oppimisosio vaaditaan seuraavan esitietona, mikä on matematiikassa tyypillistä.

Eriytetty ja oppijakeskeinen tavoiteoppiminen

Tavoiteoppimisessa on tavallisesti kaikille yhteiset ja ennalta määritellyt oppimistavoitteet, millä pyritään ennen kaikkea ryhmän sisäisten oppimiseröjen kaventamiseen. Tavoiteoppimisstrategiaa pystyy kuitenkin soveltamaan, vaikka oppimistavoitteet olisivat oppijan mukaan mukautetut.

Näin on mahdollista eriyttää opetusta, jos tavallinen opetus on oppijalle selvästi liian haastavaa tai liian helppoa. Ylöspäin eriyttämiseen LFM tosin tarjoaa jo yhtenä keinona syventäviä aktiviteetteja. Tämä on ryhmän mukaan etenevässä opetuksessa käytännöllinen, mutta ei välttämättä lahjakkaan oppijan kannalta optimaalinen tapa. Se kuitenkin palvelee hyvin niitä opiskelijoita, jotka saavuttavat tavoitetason heti joissain, mutta ei kaikissa oppimisosioissa.

Jotta yksilöidystä tavoiteoppimisstrategiasta olisi hyötyä, tavoitetason täytyy olla riittävän korkea. Ennen kaikkea yksilöllistä tavoitetasoa ei pitäisi laskea niin matalalle, että oppija aina saavuttaa tavoitetason heti.

Oppijan voi antaa myös asettaa oppimistavoitteensa itse, jos opetuksesta halutaan vähemmän opettajakeskeistä. Tällöinkin on huolehdittava siitä, että tavoitteeseen pääsyä arvioidaan riittävän luotettavalla tavalla, ja että tavoitetaso asetetaan riittävän korkealle.

3.2 Käytössä olevista arviointimenetelmistä

Tavoiteoppimisen ja tutkimuksessa esille tulleiden huomioiden valossa on sopivaa pohtia joitakin yliopistojen ja toisen asteen matematiikan opetuksessa käytössä olevia arviointikäytänteitä. Tavoiteoppiminen antaa osaltaan vihjeitä siitä, miten arviointia on mahdollista kehittää paremmin oppimista tukevaksi. Tutkimuksessa suoritettu tilastollinen analyysi taas nostaa kysymyksiä yleisesti arvioinnin luotettavuudesta.

3.2.1 Harjoitustehtävapisteeet

Tyypillisesti matematiikan kurssin summatiivinen arviointi pohjautuu kokeeseen, jota täydennetään harjoitustehtävistä saatavilla pisteillä joko niin, että ne muodostavat osan kokonaisarvioinnista (esim. Aalto-yliopiston peruskursseilla 40%) tai nostavat koepistemäärää.

Tähän on oletettavasti kaksi syytä. Ensinnäkin pisteet toimivat opiskelijoille ulkoisena motivaationa tehtävien tekoon, minkä toivotaan johtavan harjoittelun lisääntymiseen ja parempiin oppimistuloksiin. Tutkimuksessa olleen kurssipalautekyselyn perusteella pisteytystapaa säätämällä tehtävien tekoa voi mahdollisesti lisätä entisestään. Vaikeampi kysymys on missä määrin tehtävien teko edistää oppimista, mikä on sopiva harjoittelun määrä, millaiset harjoitustehtävät ovat hyödyllisimpiä ja miten opiskelijoiden erilaiset harjoittelun tarpeet saisi otettua huomioon.

Toisekseen arvioinnin *monipuolisuuden* nähdään parantavan arviointia. Esim. lukiolaki ja lukion opetussuunnitelma suorastaan velvoittavat arvioimaan oppimista "monipuolisesti", joskin tarkemmin määrittelemättä mitä monipuolisuus tarkoittaa.

Arvioinnin monipuolisuutta voi tarkastella ainakin kahdesta näkökulmasta. Arviointi

voi olla monipuolista ulkoisesti eli arviointimenetelmien osalta, tai sisällöllisesti eli kattaen erilaista ja eri aihealueiden osaamista.

Kaikkien kurssin aiheiden kattaminen kokeella ei usein ole mahdollista aikarajoitteiden vuoksi, joten tässä mielessä tehtävät voivat monipuolistaa arviointia sisällöllisesti. Tutkimuksen perusteella "tehtäväosaaminen" saattaa poiketa jossain määrin "koeosaamisesta", joten tehtävien ottaminen arviointiin mukaan voisi ajatella monipuolistavan arviointia myös tässä mielessä.

Vaikka hyväksyisi tehtävapisteen ja koepisteiden mittaavan jossain määrin erilaista osaamista, tehtäväpisteisiin liittyy tällöinkin joitakin luotettavuusongelmia. Tehtäväpisteistä yksin on esim. vaikeata päätellä, onko 0 pisteen suorituksessa kyse yrityksen vai osaamisen puutteesta. Harjoitustehtäviä tehdään myös hyvin vaihtelevissa olosuhteissa. Ei ole selvää, miten esim. avun saaminen vaikuttaa tehtäviin mittareina. On mahdollista, että saatu apu parantaa osaamista, mutta toisaalta se myös luultavasti helpottaa tehtävää.

Tehtäväpisteet saattavat siis mitata osaamisen lisäksi esim. motivaatiota, käytössä olevaa aikaa tai saadun avun määrää. Lisäksi kuten aikaisemmassa osiossa todettiin, tehtäväpisteet eivät suoraan mittaa osaamista oppimisen päätteeksi vaan oppimisen aikana. Tämän ongelmallisuus riippuu siitä, mitä summatiivisella arvioinnilla tavoitellaan. Useimmiten kuitenkin arvokkainta sekä opiskelijalle että sertifioinnin kannalta olisi tuntea juuri tämänhetkinen eikä aikaisempi osaaminen.

3.2.2 Summatiivisesta arvioinnista formatiiviseen

Matematiikan kursseilla tapahtuva muodollinen arviointi keskittyy usein summatiiviseen arviointiin. Tämän lisäksi voi tapahtua epämuodollisempaa formatiivista arviointia esim. opettajan antaessa suullista palautetta harjoittelun yhteydessä.

Muodollista formatiivista arviointia saisi kuitenkin lisättyä helposti hyödyntämällä harjoitustehtävapisteen ja kurssikokeiden kaltaisia arviointivälineitä formatiivisesti. Tämä onnistuu Bloomin tavoiteoppimisen hengessä, kun harjoitustehtävien ja kokeiden esiintuomiin osaamisen puutteisiin toteamisen lisäksi järjestelmällisesti puututaan.

Yksinkertaisimmillaan harjoitustehtävien ja kokeen uusiminen tulisi tehdä helpoksi tai jopa pakolliseksi alimpien arvosanojen kohdalla. Tässä tietokoneavusteinen arviointi on erityisen hyödyllistä, koska arviointi on välitöntä ja arviointityön määrä voisi muutoin moninkertaistua. Itse- ja vertaisarvioinnin käyttö voi myös toimia ratkaisuna silloin, kun tietokoneavusteinen arviointi ei ole teknisesti mahdollista.

Tavoitteena ei kuitenkaan ole vain mitata osaamista kahdesti, vaan että osaaminen yrityskertojen välillä myös paranisi. Tämän vuoksi ohjausta tai itseopiskelumateriaalia tulisi olla tarjolla myös ensimmäisen yrityskerran jälkeen. Kurssikokeiden kohdalla ajankohtaa voisi siirtää kurssin lopusta aiemmaksi, ja käyttää kurssin loppu ilmentyneiden

puutteiden korjaamiseen tai osaamisen syventämiseen.

Kirjallisuus

- Abacus (2016). *STEM education material bank Abacus*. URL: <https://abacus.aalto.fi/> (viitattu 25.10.2016).
- Anderson, Lorin W ja Robert B Burns (1987). "Values, evidence, and mastery learning". *Review of Educational Research* 57.2, s. 215–223.
- Anderson, Stephen A (1994). "Synthesis of Research on Mastery Learning."
- Arlin, Marshall (1984). "Time, equality, and mastery learning". *Review of Educational Research* 54.1, s. 65–86.
- Arnold, Kimberly E ja Matthew D Pistilli (2012). "Course signals at Purdue: Using learning analytics to increase student success". Teoksessa: *Proceedings of the 2nd international conference on learning analytics and knowledge*. ACM, s. 267–270.
- Baker, Eva L. (2012). "Assessment in Learning". Teoksessa: *Encyclopedia of the Sciences of Learning*. Toim. Norbert M. Seel. Boston, MA: Springer US, s. 316–321. ISBN: 978-1-4419-1428-6. DOI: 10.1007/978-1-4419-1428-6_22. URL: http://dx.doi.org/10.1007/978-1-4419-1428-6_22.
- Baker, Ryan Shaun ja Paul Salvador Inventado (2014). "Educational data mining and learning analytics". Teoksessa: *Learning analytics*. Springer, s. 61–75.
- Bergmann, Jonathan ja Aaron Sams (2012). *Flip your classroom: Reach every student in every class every day*. International Society for Technology in Education.
- Bloom, Benjamin S (1987). "A Response to Slavin's Mastery Learning Reconsidered." *Review of Educational Research* 57.4, s. 507–508.
- Bloom, Benjamin S et al. (1971). "Handbook on formative and summative evaluation of student learning."
- Bloom, Benjamin S (1976). *Human characteristics and school learning*. McGraw-Hill.
- Bloom, Benjamin S (1968). "Learning for Mastery". *Evaluation comment* 1.2, s. 1–12.
- Bloom, Benjamin S (1984). "The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring". *Educational researcher* 13.6, s. 4–16.
- Boggs, Stacey, Mark Shore ja JoAnna Shore (2004). "Using e-learning platforms for mastery learning in developmental mathematics courses". *Mathematics and Computer Education* 38.2, s. 213.

- Bull, Joanna ja Colleen McKenna (2003). *A blueprint for computer-assisted assessment*. Routledge.
- Burton, John K, David M Moore ja Susan G Magliaro (1996). "Behaviorism and instructional technology". *Handbook of research for educational communications and technology*, s. 46–73.
- Capaldi, Franco M (2014). "Mastery learning in Statics using the STEMSE online learning environment". Teoksessa: *American Society for Engineering Education (ASEE Zone 1), 2014 Zone 1 Conference of the IEEE*, s. 1–3.
- Carroll, John B (1989). "The Carroll model a 25-year retrospective and prospective view". *Educational Researcher* 18.1, s. 26–31.
- Chatti, Mohamed Amine et al. (2012). "A reference model for learning analytics". *International Journal of Technology Enhanced Learning* 4.5-6, s. 318–331.
- Ertmer, Peggy A ja Timothy J Newby (2013). "Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective". *Performance Improvement Quarterly* 26.2, s. 43–71.
- Fennell, F. (Skip), Beth Kobett ja Jonathan A. Wray (2015). "Classroom-Based Formative Assessments — Guiding Teaching and Learning". Teoksessa: *Assessment to Enhance Teaching and Learning*. Toim. Christine Suurtamm. Annual Perspectives in Mathematics Education. National Council of Teachers of Mathematics, s. 51–62.
- Fosnot, Catherine Twomey ja Randall Stewart Perry (1996). "Constructivism: A psychological theory of learning". *Constructivism: Theory, perspectives, and practice* 2, s. 8–33.
- Gagne, Robert (1985). *The Conditions of Learning and Theory of Instruction* Robert Gagné. New York, NY: Holt, Rinehart ja Winston.
- Grant, Lyle K ja Robert E Spencer (2003). "The personalized system of instruction: Review and applications to distance education". *The International Review of Research in Open and Distributed Learning* 4.2.
- Guskey, Thomas R (1986). "Defining the Critical Elements of a Mastery Learning Program."
- Guskey, Thomas R (1985). *Implementing mastery learning*. Wadsworth Publishing Company.
- Guskey, Thomas R (2008). "Mastery Learning". Teoksessa: *21st century education: A reference handbook*. Toim. Thomas L Good. Sage Publications, s. 194–202.
- Guskey, Thomas R (1987). "Rethinking mastery learning reconsidered". *Review of Educational Research* 57.2, s. 225–229.
- Guskey, Thomas R ja Sally L Gates (1986). "Synthesis of research on the effects of mastery learning in elementary and secondary classrooms". *Educational Leadership* 43.8, s. 73–80.

- Hoon, Teoh Sian, Toh Seong Chong ja Nor Azilah Binti Ngah (2010). "Effect of an Interactive Courseware in the Learning of Matrices." *Educational Technology & Society* 13.1, s. 121–132.
- Horton, Lowell (1979). "Mastery Learning: Sound in Theory, But..." *Educational Leadership* 37.2, s. 154–56.
- Jungic, Veselin, Petra Menz ja Randall Pyke (2011). *A Collection of Problems in Differential Calculus*. URL: <http://faculty.ung.edu/jallagan/Courses%20materials/Math%201450%20Calculus%201/Syllabus%20and%20ebook/problems%20and%20solutions%20for%20calculus%201.pdf> (viitattu 03.05.2017).
- Knight, Peter T (2002). "Summative assessment in higher education: practices in disarray". *Studies in higher Education* 27.3, s. 275–286.
- Kulik, Chen-Lin C, James A Kulik ja Robert L Bangert-Drowns (1990). "Effectiveness of mastery learning programs: A meta-analysis". *Review of educational research* 60.2, s. 265–299.
- Lai, Patrick ja John Biggs (1994). "Who benefits from mastery learning?" *Contemporary Educational Psychology* 19.1, s. 13–23.
- Leonard, William J, CV Holot ja William J Gerace (2008). "Mastering Circuit Analysis: An innovative approach to a foundational sequence". Teoksessa: *2008 38th Annual Frontiers in Education Conference*. IEEE, F2H–3.
- Matta (2017). *Matta-tutkimusryhmän julkaisuluettelo*. URL: <https://math.aalto.fi/fi/opiskelu/matta/tutkimus/> (viitattu 24.11.2017).
- McGaghie, William C, S Barry Issenberg et al. (2011). "Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence". *Academic medicine: journal of the Association of American Medical Colleges* 86.6, s. 706.
- McGaghie, William C, Saul B Issenberg et al. (2014). "A critical review of simulation-based mastery learning with translational outcomes". *Medical Education* 48.4, s. 375–385.
- Miller, Allen Horace, Bradford William Imrie ja Kevin Cox (1998). *Student Assessment in Higher Education: A Handbook for Assessing Performance*. Psychology Press.
- Moodle (2017). *Analytics - MoodleDocs*. URL: <https://docs.moodle.org/34/en/Analytics> (viitattu 24.11.2017).
- Morgan, Kelly (2011). *Mastery Learning in the science classroom: Success for every student*. NSTA Press.
- Motamedi, Vahid ja William J Sumrall (2000). "Mastery learning and contemporary issues in education". *Action in Teacher Education* 22.1, s. 32–42.
- Mueller, Daniel J (1976). "Mastery learning: Partly boon, partly boondoggle." *Teachers College Record*.

- Nohda, Nobuhiko (2000). "Teaching by Open-Approach Method in Japanese Mathematics Classroom."
- Page, G Terry, John Bernard Thomas ja Alan R Marshall (1979). "International dictionary of education".
- Pelkola, Timo, Antti Rasila ja Christopher Sangwin (2018). "Investigating Bloom's Learning for Mastery in Mathematics with Online Assessment." *Informatics in Education* 17.2, s. 363–380.
- Phillips, Denis C. (2012). "Behaviorism and Behaviorist Learning Theories". Teoksessa: *Encyclopedia of the Sciences of Learning*. Toim. Norbert M. Seel. Springer US, s. 438–442. ISBN: 978-1-4419-1428-6.
- Rasila, Antti (2016). "E-Assessment Material Bank Abacus". *To appear EDULEARN16 Proceedings*.
- Roediger III, Henry L ja Elizabeth J Marsh (2005). "The positive and negative consequences of multiple-choice testing." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31.5, s. 1155.
- Rohrer, Doug ja Kelli Taylor (2007). "The shuffling of mathematics problems improves learning". *Instructional Science* 35.6, s. 481–498.
- Sangwin, Chris (2013). *Computer aided assessment of mathematics*. OUP Oxford.
- Schunk, Dale H (1996). "Learning theories". *Printice Hall Inc., New Jersey*.
- Scouller, Karen (1998). "The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay". *Higher Education* 35.4, s. 453–472.
- Sherriffs, Alex C ja Donald S Boomer (1954). "Who is penalized by the penalty for guessing?" *Journal of Educational Psychology* 45.2, s. 81.
- Siemens, George ja Ryan SJ d Baker (2012). "Learning analytics and educational data mining: towards communication and collaboration". Teoksessa: *Proceedings of the 2nd international conference on learning analytics and knowledge*. ACM, s. 252–254.
- Slavin, Robert E (1987). "Mastery learning reconsidered". *Review of educational research* 57.2, s. 175–213.
- Smith, Geoff et al. (1996). "Constructing mathematical examinations to assess a range of knowledge and skills". *International Journal of Mathematical Education in Science and Technology* 27.1, s. 65–77.
- Taras, Maddalena (2005). "Assessment–summative and formative–some theoretical reflections". *British Journal of Educational Studies* 53.4, s. 466–478.
- Tiitu, Hannu (2016). "Anna Karenina-periaate. Kohti matematiikan oppimisympäristössä toteutettua virheiden luokittelua". Tutkielma. University of Helsinki.
- Toivanen, Aurora (2013). "Yksilöllisen oppimisen malli Martinlaakson lukion matematiikan opetuksessa". Tutkielma. University of Helsinki.

- Trenholm, Sven, Lara Alcock ja Carol Robinson (2015). "An investigation of assessment and feedback practices in fully asynchronous online undergraduate mathematics courses". *International Journal of Mathematical Education in Science and Technology* 46.8, s. 1197–1221.
- Yeo, Joseph BW (2007). "Mathematical tasks: Clarification, classification and choice of suitable tasks for different types of learning and assessment".
- Ylioppilastutkintolautakunta (2016). *Matematiikan koe, pitkä oppimäärä 28.9.2016 Hyvän vastauksen piirteitä*. URL: https://www.ylioppilastutkinto.fi/images/sivuston_tiedostot/Hyv_vast_piirt/FI_2016_S/2016_S_M.pdf (viitattu 17.10.2016).
- Zhao, Yuyuan (2006). "How to design and interpret a multiple-choice-question test: a probabilistic approach". *International Journal of Engineering Education* 22.6, s. 1281.